Vol. 33, No. 6 December, 1949

Journal of Applied Psychology

Edited By: Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation;* Walter V. Bingham, *Washington, D.C.;* Harold E. Burtt, *Ohio State University;* Allen L. Edwards, *University of Washington;* Irving Lorge, *T. C. Columbia University;* Quinn McNemar, *Stanford University;* James P. Porter, *Danville, Illinois;* Julian B. Rotter, *Ohio State University;* Edward K. Strong, Jr., *Stanford University;* Donald E. Super, *T. C. Columbia University;* Morris S. Viteles, *University of Pennsylvania;* Alfred C. Welch, *Knox-Reeves, Minneapolis.*

Table of Contents

Study of Executive Leadership in Business: I. The R, A, and D Scales: C. G. Browne 521
An Analysis of the Leaderless Group Discussion: B. M. Bass 527
Performance on the File-Remmers Test, How Supervise? Before and After a Course in Psychology. H. W. Karn 534
The Prediction of Accidents of Taxicab Drivers: E. E. Ghiselli and C. W. Brown 540
Some Precautions in the Use of the Per Cent Method of Job Evaluation: W. D. Turner 547
Predicting Subject Grades of Liberal Arts Freshmen with the Kuder Preference Record: D. T. Hake and C. H. Kuebisli 553
MMPI Personality Patterns for Various Occupations: E. E. Daniels and W. A. Hunter 559
Correcting Special Ability Test Scores for General Ability: A. Levine 566
The Rorschach Test in Industrial Selection: A. F. Rieger 569
The Rorschach Test and Occupational Personalities: A. F. Rieger 572
A Factorial Analysis of Arm-Hand Precision Tests: R. H. Seashore, F. J. Dudek and W. Holtzman 579
What Do Readership Studies Really Prove? H. P. Longstaff and G. P. Laybourn 585
Psychological Factors in Instrument Reading: II. The Accuracy of Pointer Position Interpolation as a Function of the Distance between Scale Marks and Illumination: W. F. Grether and A. C. Williams, Jr. 594
Identification of Cola Beverages. III. A Final Study: N. H. Pronko and J. W. Bowles, Jr. 605
Book Reviews 609
New Books, Monographs, and Pamphlets 615

Published Bi-monthly by The American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa., and
1515 Massachusetts Ave., NW, Washington 5, D. C.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. and R., authorized October 10, 1947

Copyright, 1949, by The American Psychological Association, Inc.

Journal of Applied Psychology

The *Journal of Applied Psychology* is published bi-monthly by the American Psychological Association, Inc. Annual subscription, $3.00, single copies, $1.25. Subscriptions, orders, and other business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Avenue, N. W., Washington 5, D. C. Articles for publication should be sent to the Editor, Donald G. Paterson, Department of Psychology, University of Minnesota, Minneapolis 14, Minn.

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 16,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 80 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described by J. E. Anderson and W. L. Valentine in "The preparation of articles for publication in the journals of the American Psychological Association," *Psychol. Bull.*, 1944, 41, 345–376. A reprint of this article will be loaned to any prospective contributor who does not find it in his library.

Journal of Applied Psychology

Edited by

Donald G. Paterson

University of Minnesota

Consulting Editors

GEORGE K. BENNETT, *Psychological Corporation;* WALTER V. BINGHAM, *Washington, D. C.;* HAROLD E. BURTT, *Ohio State University;* ALLEN L. EDWARDS, *University of Washington;* IRVING LORGE, *T. C. Columbia University;* QUINN MCNEMAR, *Stanford University;* JAMES P. PORTER, *Danville, Illinois;* JULIAN B. ROTTER, *Ohio State University;* EDWARD K. STRONG, JR., *Stanford University;* DONALD E. SUPER, *T. C. Columbia University;* MORRIS S. VITELES, *University of Pennsylvania;* ALFRED C. WELCH, *Knox-Reeves, Minneapolis.*

Volume 33, 1949

Published Bi-monthly by The American Psychological Association, Inc.

Prince and Lemon Sts., Lancaster, Pa., and 1515 Massachusetts Ave., NW, Washington 5, D. C.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, Section 538, P. L. & R., authorized October 10, 1947

Copyright, 1949, by The American Psychological Association, Inc.

LANCASTER PRESS, INC., LANCASTER, PA.

Contents of Volume 33

Articles

Aldrich, M. G. **A Follow-up Study of Social Guidance at the College Level** 258

Anderson, R. G. **Reported and Demonstrated Values of Vocational Counseling** 460

Angoff, W. H. **An Empirical Approach to a Problem of Psychophysical Scaling** 59

Barnett, A. **A Note on Mechanical Aptitude of West Texans** 316

Bass, B. M. **An Analysis of the Leaderless Group Discussion** 527

Browne, C. G. **Study of Executive Leadership in Business. I. The R, A, and D Scales** 521

Carrington, D. H. **Note on the Cardall Practical Judgment Test** 29

Chesler, D. J. **Abbreviated Job Evaluation Scales Developed on the Basis of "Internal" and "External" Criteria** 151

Clark, K. E. **A Vocational Interest Test at the Skilled Trades Level** 291

Daniels, E. E. and Hunter, W. A. **MMPI Personality Patterns for Various Occupations** 559

DiMichael, S. G. **Work Satisfaction and Work Efficiency of Vocational Counselors as Related to Measured Interests** 319

Donceel, J. F., Alimena, B. S. and Birch, C. M. **Influence of Prestige Suggestion on the Answers of a Personality Inventory** 352

Dougan, C. P., Schiff, E. and Welch, L. **Originality Ratings of Department Store Display Department Personnel** 31

Edwards, A. S. **Attention and Involuntary Movement** 503

Elinson, J. **Attitude Research in the Army** 1

Farr, J. N. and Jenkins, J. J. **Tables for Use with the Flesch Readability Formulas** 275

Fehrer, E. and Strupp, H. **The Effect of Equating Interest Test Items for Prestige Value** 222

Feronte, N. C. **Tests Used by United States Air Carriers** 445

Ford, A. **Types of Errors in Location Judgments on Scaled Surfaces. I. Errors of Configuration** 373

Ford, A. **Types of Errors in Location Judgments on Scaled Surfaces. II. Random and Systematic Errors** 382

Ghiselli, E. E. and Brown, C. W. **The Prediction of Accidents of Taxicab Drivers** 540

Giese, W. J. and Ruter, H. W. **An Objective Analysis of Morale**.... 421
Glanz, E. **A Trade Test for Power Sewing Machine Operators**..... 436
Gordon, T. **The Airline Pilot's Job**..... 122
Greene, J. E., Osborne, R. T. and Sanders, W. B. **A Window-Stencil Method for Scoring the Strong Vocational Interest Blank (Men)**..... 141
Grether, W. F. **Instrument Reading. I. The Design of Long-Scale Indicators for Speed and Accuracy of Quantitative Readings**..... 363
Grether, W. F. and Williams, A. C., Jr. **Psychological Factors in Instrument Reading. II. The Accuracy of Pointer Position Interpolation as a Function of the Distance between Scale Marks and Illumination**..... 594
Hadley, J. M. and Kahn, D. F. **A Comment on Wallace's Note on "Factors Related to Life Insurance Selling"**..... 359
Hake, D. T. and Ruedisili, C. H. **Predicting Subject Grades of Liberal Arts Freshmen with the Kuder Preference Record**..... 553
Harrell, T. W., Brown, D. E. and Schramm, W. **Memory in Radio News Listening**..... 265
Harris, F. J. **The Quantification of an Industrial Employee Survey. I. Method**..... 103
Harris, F. J. **The Quantification of an Industrial Employee Survey. II. Application**..... 112
Holdrege, F. E., Jr. **Implementing an Employee Opinion Survey**.. 428
Jaspen, N. **A Factor Study of Worker Characteristics**..... 449
Jenkins, W. L. and Connor, M. B. **Some Design Factors in Making Settings on a Linear Scale**..... 395
Jurgensen, C. E. **A Fallacy in the Use of Median Scale Values in Employee Check Lists**..... 56
Kahn, D. F. and Hadley, J. M. **Factors Related to Life Insurance Selling**..... 132
Karn, H. W. **Performance on the File-Remmers Test, How Supervise? Before and After a Course in Psychology**..... 534
Katz, D. **An Analysis of the 1948 Polling Predictions**..... 15
Kerr, W. A. and Martin, H. L. **Prediction of Job Success from the Application Blank**..... 442
Kirchheimer, B. A., Axelrod, D. W. and Hickerson, G. X., Jr. **An Objective Evaluation of Counseling**..... 249
Kirkpatrick, J. J. and Cureton, E. E. **Vocabulary Item Difficulty and Word Frequency**..... 347
Knauft, E. B. **A Selection Battery for Bake Shop Managers**..... 304
Kriedt, P. H. **Vocational Interests of Psychologists**..... 482

Kriedt, P. H. and Clark, K. E. **"Item Analysis" Versus "Scale Analysis"** 114
Kuntz, J. E. and Sleight, R. B. **Effect of Target Brightness on "Normal" and "Subnormal" Visual Acuity** 83
Lawshe, C. H. and Farbro, P. C. **Studies in Job Evaluation: 8. The Reliability of an Abbreviated Job Evaluation System** 158
Lawshe, C. H., Kephart, N. C. and McCormick, E. J. **The Paired Comparison Technique for Rating Performance of Industrial Employees** 69
Levine, A. S. **Correcting Special Ability Test Scores for General Ability** 566
Link, H. C. and Freiberg, A. D. **The Psychological Barometer on Communism, Americanism and Socialism** 6
Locke, B. and Grimm, C. H. **Odor Selection, Preferences and Identification** 167
Longstaff, H. P. and Laybourn, G. P. **What Do Readership Studies Really Prove?** 585
Lyman, H. B. **Flesch Count and Readership of Articles in a Midwestern Farm Paper** 78
McCandless, B. R. **The Rorschach as a Predictor of Academic Success** 43
Mintz, A. and Blum, M. L. **A Re-examination of the Accident Proneness Concept** 195
Mosier, M. F. and Kuder, G. F. **Personal Preference Differences Among Occupational Groups** 231
Ostrom, S. R. **The OL Key of the Strong Test and Drive at the Twelfth Grade Level** 240
Ostrom, S. R. **The OL Key of the Strong Vocational Interest Blank for Men and Scholastic Success at College Freshmen Level** 51
Otis, J. L. and Chesler, D. J. **A Short Test of Mental Ability** 146
Perloff, E. **Prediction of Female Readership of Magazine Articles** .. 175
Pronko, N. H. and Bowles, J. W., Jr. **Identification of Cola Beverages. III. A Final Study** 605
Rieger, A. F. **The Rorschach Test and Occupational Personalities** .. 572
Rieger, A. F. **The Rorschach Test in Industrial Selection** 569
Satter, G. A. **Method of Paired Comparisons and a Specification Scoring Key in the Evaluation of Jobs** 212
Seashore, R. H., Dudek, F. J. and Holtzman, W. **A Factorial Analysis of Arm-Hand Precision Tests** 579
Shaffer, R. H. **Kuder Interest Patterns of University Business School Seniors** 489
Sherriffs, A. C. **Modification of Academic Performance Through Personal Interview** 339

Sinaiko, H. W. **The Rosenzweig Picture-Frustration Study in the Selection of Department Store Section Managers** 36
Strong, E. K., Jr. **Vocational Interests of Accountants** 474
Thorndike, E. L. **Note on the Shifts of Interest with Age** 55
Tiffin, J., Parker, B. T. and Habersat, R. W. **Visual Performance and Accident Frequency** 499
Tinker, M. A. and Paterson, D. G. **Speed of Reading Nine Point Type in Relation to Line Width and Leading** 81
Turner, W. D. **Some Precautions in the Use of the Per Cent Method of Job Evaluation** 547
Wallace, S. R., Jr. **A Note on Kahn and Hadley's "Factors Related to Life Insurance Selling"** 356
Whitlock, J. B. and Crannell, C. W. **An Analysis of Certain Factors in Serious Accidents in a Large Steel Plant** 494
Wittenborn, J. R. **Certain Rorschach Response Categories and Mental Abilities** 330

Book Reviews

Achilles' Management and the Psychologist: A Practical Guide on Psychology for the Business Executive: Donald G. Paterson 187
Ahern's Survey of Personnel Practices in Unionized Offices: C. E. Jurgensen 187
Bowler and Dawson's Counseling Employees: C. E. Jurgensen 279
Buros' The Third Mental Measurements Yearbook: E. Donald Sisson 181
Burtt's Applied Psychology: Steuart Henderson Britt 510
Chapin's Experimental Designs in Sociological Research: Donald E. Super 93
Clarke's The Application of Measurement to Health and Physical Education: C. H. McCloy 98
Darley's The Use of Tests in College and Froechlich and Benson's Guidance Testing: Milton E. Hahn 96
deFord's Psychologist Unretired, the Life Pattern of Lillian Martin: Carl E. Seashore 612
Doob's Public Opinion and Propaganda: Alfred C. Welch 284
Erickson's A Basic Text for Guidance Workers: William A. McClelland 97
Escalona's An Application of the Level of Aspiration Experiment to the Study of Personality: Julian B. Rotter 515
Evans' An Introduction to Color: Miles A. Tinker 416
Eysenck's Dimensions of Personality: Arthur Weider 614
Goldstein's The Roots of Prejudice Against the Negro in the United States: Allen L. Edwards 516

Jucius' Personnel Management: Albert S. Thompson 414
Kaback's Vocational Personalities: An Application of the Rorschach Group Method: Boyd McCandless 612
Kaufmann's Your Job: William A. McClelland 517
Kessler's Rehabilitation of the Physically Handicapped: Donald H. Dabelstein 279
Lall's Mental Measurement: Henry E. Garrett 415
Lawshe's Principles of Personnel Testing: Edwin E. Ghiselli 92
Lewin's Resolving Social Conflicts: Horace B. English 410
Linebarger's Psychological Warfare: Clark L. Hosmer 187
OSS Assessment Staff's Assessment of Men: Selection of Personnel for the Office of Strategic Services: Donald E. Super 511
Pigors and Myers' Personnel Administration: A Point of View and a Method: Albert S. Thompson 282
Planty, McCord and Efferson's Training Employees and Managers for Production and Teamwork: Clifford E. Jurgensen 611
Ross's Measurement in Today's Schools and Ross's Chapter Exercises and Tests to Accompany Measurement in Today's Schools: Walter W. Cook 99
Rudolph's Attention and Interest Factors in Advertising: Howard P. Longstaff 286
Selekman's Labor Relations and Human Relations: Brent Baxter 287
Stouffer, et al. The American Soldier: Volume I, Adjustment During Army Life; Volume II, Combat and its Aftermath: Allen L. Edwards 609
Terman and Oden's The Gifted Child Grows Up: Twenty-five Years' Follow-up of a Superior Group: Sidney L. Pressey 189
Yoder's Personnel Management and Industrial Relations: Albert S. Thompson 280
Yoder, Paterson, et al. Local Labor Market Research: Arthur H. Brayfield 411

Miscellaneous

New Books, Monographs, and Pamphlets. 101, 191, 289, 418, 519, 615
Erratum 100

Journal of Applied Psychology

VOL. 33, NO. 6 DECEMBER, 1949

Study of Executive Leadership in Business. I. The R, A, and D Scales

C. G. Browne*

Wayne University

This is the first in a series of papers which will present the following methods for the study of leadership and executive relationships in business: R, A, and D scales; social and organizational contacts; sociometric pattern; Goal and Achievement index (1).

The total study proceeded on the following hypotheses: (1) leadership is a process based upon the inter-relationships of individuals in a group which is working toward a goal that has been accepted as desirable; (2) executive function and leadership in business is a process of the interaction of social and working relationships within and outside of the executive groups; and (3) executive and leader relationships can be analyzed through the application of methods which are not designed to measure personal executive traits as psychological entities.

Procedure

The subjects in these explorations were 24 executives of a tire and rubber company in Ohio, named the Congo Tire and Rubber Company for purposes of the study. Table 1 includes a listing of the executives by title and department. All of the company executives on the first, second, and fourth echelons of the business, and all of the executives on the third echelon with one exception were included. Data were obtained in a moderately structured interview, varying in length from 2½ to 3½ hours. Some of the executives completed the R, A, and D scales during the interview, while others completed them at another time. In all cases, the scales were explained during the interview.

R, A, and D Scales

The RAD index form devised by Stogdill and Shartle in their studies of Naval leadership consists of six scales, each containing eight state-

* The writer is indebted to Drs. C. L. Shartle, Harold E. Burtt, and Ralph M. Stogdill of the Ohio State University for their guidance and criticisms throughout the study.

ments (4). Scales A and B are for Responsibility; scales C and D, for Authority; and scales E and F, for Delegation of authority. The person completing the forms checks his first and second choices of statements as they best apply to him on each of the six scales.[1] The following are examples of the statements for each of the variables: *Responsibility*, "I am responsible for the successful operation and coordination of all activities in the organization"; *Authority*, "I make no decisions whatsoever but request instructions from my superior on all matters"; *Delegation of authority*, "I have delegated full authority to my assistants, allowing them complete right of decision in all functions."

Scoring of the individual items on each scale was developed using the Thurstone equal appearing interval technique (5). To establish scale values, the statements were evaluated by staff, graduate students, and seniors in psychology at the Ohio State University. The mean of the point values of the four statements checked is the score on that variable. Scale values for the statements range from 1.0 (indicating a high degree of the factor) to 8.7 (indicating a low degree). *Therefore, the lower scores indicate a higher degree of the item measured, while the higher scores indicate a lower degree.*

R, A, and D Scores

The R, A, and D scores of each executive, the mean scores by departmental and total groups, and the range for each factor are given in Table 1. Remembering that the lower scores indicate a higher degree of the factor, the score of 1.6 for the president and general manager represents the highest for both R and A. Likewise, the scores of 5.2 and 5.1 for the manager of tube sales represent the lowest for R and A, respectively. The vice-president-sales had the highest D score, and he was also one of the three executives who received the greatest number of choices on the sociometric diagram. While the secretary of the company had the lowest D score (6.7), an analysis of his work revealed that he had no one under his supervision to whom he could delegate the relatively small degree of authority which he estimated he had.

In any measure—individual scores, departmental means, or total means—the R scores were almost consistently the highest, followed by the A scores, and finally the D scores. This indicates a general trend for the executives to estimate that they delegated authority in a lesser amount than they estimated either their responsibility or authority, and that their authority was less than their responsibility. Although the ranges of the R and A scores were almost identical, there was a concentration of R scores, there being 22 cases between 2.7 and 3.9, with a mode of

[1] Requests for information regarding the R, A, and D scales may be addressed to Dr. R. M. Stogdill, Associate Director, Ohio State University Leadership Studies, The Ohio State University, Columbus, 10, Ohio.

Table 1

R, A, and D Scores

Executive Department and Title	R Score	A Score	D Score
General Administration			
President and genl. manager	1.6	1.6	2.7
Secretary of the company	3.2	4.3	6.7
Director public relations	3.3	2.9	3.8
Purchasing agent	3.9	3.6	6.4
Department mean	3.0	3.1	4.9
Sales			
Vice-president-sales	2.7	3.5	2.3
Sales manager	2.7	2.9	3.0
Manager Congo stores	2.7	3.8	4.5
Manager sales promotion	2.7	3.4	4.9
Manager sales orders	2.7	4.4	4.9
Manager tube sales	5.2	5.1	5.1
Department mean	3.1	3.8	4.1
Finance			
Treasurer	2.7	4.1	4.9
Comptroller	2.7	3.4	3.8
Supervisor cost accounting	2.9	4.1	5.2
Chief accountant	3.3	3.4	6.2
Department mean	2.9	3.7	5.0
Manufacturing			
Vice-president-manufacturing	3.0	2.9	3.2
Plant engineer	3.7	4.4	4.0
Chief chemist	3.7	3.4	3.8
Product engineer	2.7	3.6	3.8
Foreman bicycle tire production	3.7	4.7	2.9
Manager production control	2.9	3.4	4.4
Manager quality control	2.7	4.5	5.5
Manager shipping	3.3	4.3	6.0
Department mean	3.2	3.9	4.2
Personnel			
Personnel director	3.2	2.3	2.5
Industrial engineer	3.2	2.3	3.3
Department mean	3.2	2.3	2.9
Total group mean	3.1	3.6	4.3
Range	1.6–5.2	1.6–5.1	2.3–6.7

2.8 for all R scores. The A scores, however, distributed more uniformly, with a mode of 4.3, while the D scores had the greatest range, but were distributed most uniformly. There were 17 D scores lower than 3.3, compared with 18 R scores of 3.3 or higher.

While these scores cannot be considered to be predictive of executives in other companies on the basis of work done, or departmental assignment, or echelon level, the method offers opportunities to study working

relationships between executives which may be related to any of these variables. As a measure of communication within the company and of other personal relationships, the R, A, and D scales offer further possibilities. These measures might be obtained by a study in which an executive's seniors complete R and A forms for him and his juniors complete D forms for him. A comparison of these scores with the executive's own forms would constitute a measure of the individual's understanding of his responsibility and authority from the seniors who determine them and of his delegation of authority from the juniors to whom delegation is made.

R, A, and D Relationships

Table 2 includes correlation coefficients between R, A, and D and other variables used in the study. *These correlations are descriptive only of the relationships existing between the variables for this particular population of executives. They cannot be interpreted as sampling statistics, since the group of executives studied here does not constitute a statistical sample.*

Table 2

Product Moment Inter-correlations of R, A, and D Scores and R, A, and D Scores Correlated with Other Variables

Variable (N = 24)	R	A	D
R (Responsibility)	x	.56	.29
A (Authority)	.56	x	.54
D (Delegation of authority)	.29	.54	x
Time spent in supervision**	−.06*	−.25*	−.12*
Number of choices***	.29*	.28*	.48*
Executive's salary	.48*	.41*	.49*
Executive's echelon	.34	.40	.14

* The sign for this correlation has been changed so that in interpreting the correlations a large score in one variable is also indicative of a large score in or a greater degree of the second variable.

** This variable was expressed in the per cent of the executive's total time which he estimated he spent in supervision.

*** As determined from the sociometric diagram.

The inter-correlations between the three factors were .56 for R and A; .29 for R and D; and .54 for A and D. In the studies of Naval leadership, unpublished correlations for a group of 40 Naval officers were found to be .56 for R and A; .16 for R and D; and .86 for A and D. These comparative correlations between the business executives and the Naval officers indicate the same general trend in the inter-correlations between factors, although the Navy correlation of .86 for A and D was considerably higher than the executive correlation of .54 for the same factors.

Some of this difference may have been due to the possibility that such concepts as authority and the delegation of it are more clearly defined for military personnel than they are for business executives, and that they are measured and weighed with greater absoluteness in the military environment.

The correlation between authority and time spent in supervision, the largest of the three negative correlations between these variables, was −.25. This indicates that the executive who devoted a greater percentage of his time to supervisory activities, as contrasted with such other activities as planning or coordination or evaluation, tended to have an A score which represented a lesser amount of authority.

In a later paper, the sociometric pattern which was used in the study will be discussed. The "number of choices" variable was determined from the listing which each executive made of the men with whom he spent most time in getting his work done. In the sociometric diagram, the greatest number of choices was received by three executives who were in the second echelon. The relatively high positive correlation of .48 between number of choices and D score indicates that those men who were consulted most and with whom most time was spent in getting work done also tended to be the men who were delegating authority to the greatest degree.

The correlations between R, A, and D scores and salary all indicate that executives with the higher salaries tended to have scores which indicated a greater degree of the three factors. In view of the low correlation between echelon and D score, the relatively high correlation of .49 between salary and D score may be surprising. However, this correlation is strongly influenced by the fact that several of the executives on the fourth echelon were receiving higher salaries than some of the executives on the second and third echelon.

The correlations between R, A, and D scores and echelon were not as high as they were with salary. However, it is quite logical that the correlations with echelon were highest for responsibility and authority, since it can be expected that the higher level executives would have a higher index on these factors. Delegation of authority, on the other hand, is an individualized factor, not greatly related to the executive's echelon.

Summary

The R, A, and D scales introduced by Stogdill and Shartle in their studies of Naval leadership were applied to a group of 24 executives in a tire and rubber manufacturing company. The scores for each executive on each of the three factors provide a measure of the individual's evaluation of his responsibility, authority, and delegation of authority. From

their scores, these executives estimated that their responsibility and authority were greater than their delegation of authority.

Since the factors measured by the R, A, and D scales are particularly important at the executive level, a quantitative method such as presented here should aid in the analysis and understanding of executive functioning and business leadership. This is based upon the general hypothesis that leadership and executive activity are dependent upon social and working relationships in group activities, and that their study from this approach will prove more helpful than the analysis of individual characteristics with psychological trait testing methods has proven.

Received March 14, 1949.

References

1. Browne, C. G. *An exploration into the use of certain methods for the study of executive function in business.* Unpublished Ph.D. dissertation, The Ohio State University, 1948.
2. Jenkins, W. O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, **44**, 54–79.
3. Stogdill, R. M. Personal factors associated with leadership—a survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.
4. Stogdill, R. M., and Shartle, C. L. Methods for determining patterns of leadership behavior in relation to organization structure and objectives. *J. appl. Psychol.*, 1948, **32**, 286–291.
5. Thurstone, L. L., and Chave, E. J. *The measurement of attitude.* Chicago: Univ. of Chicago Press, 1929.

An Analysis of the Leaderless Group Discussion

Bernard M. Bass

Louisiana State University

Several studies in recent years have reported the use of the leaderless group discussion situation as an aid in selecting candidates for positions involving leadership (**1, 2, 3, 4**). Little has been done to evaluate this technique quantitatively or to investigate the possibility of making objective measures of individuals in this situation. The purposes of the present study were to investigate the extent of agreement among raters of discussion participants and the relation between the total amount of time a participant spent talking[1] in the leaderless group discussion, and the ratings he obtained.

Subjects, Method and Apparatus

A class of 20 educational psychology students served as subjects. Twelve were men and 8 were women. They ranged from freshmen to graduate student, with a median of 2 years college education. Several students had 2 or more years of teaching experience.

A total of 6 leaderless group discussions was run in 6 weeks. The 20 subjects were divided randomly into Group A and Group B. Group A participated in the first discussion while Group B observed the participants. The two groups of 10 students each switched roles for the second discussion. Those 10 participants of the first 2 discussions who had been given the highest leadership ratings for the discussions by their classmates formed the third leaderless group discussion. The fourth discussion was composed of the remaining 10 participants who had received the lowest leadership ratings. The original groups, A and B, were used again for the last 2 discussions. Group A participated in and Group B observed the fifth discussion, and Group B participated in and Group A observed the sixth discussion.

Each of the discussions lasted 20 minutes and was held during class hours in the classroom. The 10 participants were seated around the outside of a V-shaped table. A code numbered place card was put in front of each of the participants for identification. The 10 observers sat facing the participants on the other side of the room. To provide adequate

[1] For a discussion of the use of time spent talking in the individual interview as a predictive measure, the reader may refer to Chapple, E. D. and Donald, G. A method of evaluating supervisory personnel. *Harv. Bus. Rev.*, 1946, **24**, 197–214.

motivation for the participants, each of the discussions were considered a course examination and grades from A to E were awarded by the experimenter-instructor.

Oral instructions were as follows:

"You will be given a problem and will have thirty minutes in which to discuss it. You will be graded, not only on how well you as an individual contribute to the group discussion, but also on how well the group does as a whole.

"Everyone may receive an A or everyone may receive an E depending on how much he contributes and how much the group progresses. Therefore, if you feel someone else is 'off the track,' is wasting the group's time and therefore is lowering your grade, feel free to cut in and get the group back on its proper assignment."

The problem presented to a discussion group was pertinent to the material on educational psychology which the subjects had supposedly studied the night before. It called for the development of a program or set of plans which must be sold to another group, such as a school board. The problem was read aloud to the group twice and then they were told to begin discussing it.

One example of the problems used is as follows:

The School Board of your town has gone progressive. The Board realizes that teachers cannot do everything and are planning to obtain a staff of specialists in various areas to cope with the several problems which teachers are unable to handle adequately. Consider yourselves as the chairmen of the ten departments of your high school of 5,000 students. You are meeting thirty minutes before the School Board goes into session. The present high school personnel consists of teachers, the principal, an office staff, and a janitorial staff. Your problem is to agree upon the four specialists you will ask for, and the reasons you will present for choosing those four. The School Board will only appropriate $12,000. And remember, there are 5,000 students, so don't plan on overloading the four specialists.

For the first 3 discussions, the experimenter used a stop-watch to clock the number of seconds each participant talked and then recorded the measurement under the participant's name on a log sheet.[2] It was found difficult to keep up with the constant shift in speakers, especially during arguments. Interruptions and pauses within speeches also increased the difficulty of measuring and recording. Fortunately, the differences among participants were large enough to allow one to assume that the measurement errors were not of such an order as to warrant discarding the results of the first 3 discussions.

In order to increase the time data precision, the Group Discussion Chronometer was designed by the investigator and introduced into the study in the fourth discussion. The GDC consisted of a panel of 10

[2] Wire recordings were not used because of the difficulties of dubbing in the speakers' code numbers which would have been necessary in order to later identify who was speaking at a given time.

button switches spaced to allow each of the experimenter's fingers to operate 1 button without having to move his hands. Each push-button closed the circuit of 1 of 10 sweep-hand, self-starting, electric clocks mounted on a board outside of the classroom and connected by a cable to the switch panel. One button and 1 clock were devoted to each participant and a cumulative measure, in seconds, was obtained of the total time each participant talked. The experimenter pressed the button appropriate to the participant each time the participant began talking and released it when the speaker stopped or paused. Unlike the stop-watch, the GDC could record individual times even when two participants attempted to talk at the same time.

Both participants and observers rated participants by 3 different rating methods on 13 items. The 3 rating methods were the "Spread N," which was used for all 6 discussions; the "3 Best minus 3 Worst," which was used for the third discussion, and the "rank order of merit," which was used for the fourth discussion. All ratings were recorded on a prepared inventory by the raters immediately after the conclusion of a discussion.

Where 10 participants were to be rated, the "Spread N" technique[3] was as follows:

Each of the 20 observers and participants assigned 10 votes on a given item to the participants in any way desired. If the rater thought one person completely outclassed all the other 9 on a given attribute, the rater assigned all 10 votes to the 1 individual and none to the others. The rater gave 1 vote to each of the 10 participants if all participants were judged equal. Any other distribution of the 10 votes was possible.

When the rank order of merit method was used, the raters were instructed to rank the 10 participants on each of the 13 items. For the "3 Best minus 3 Worst" method, the raters were told to select the 3 highest and 3 lowest participants on each of the items.

The 13 items upon which appraisals were made were as follows:

Vote for the person or persons:

1. Whom you think led the discussion.
2. Whom you think knew most about the topic discussed.
3. Whom you think most influenced the other participants in the discussion.
4. Who most clearly defined the problems, who brought them into sharp focus, and who best organized the group's thinking during the discussion.
5. Whom you would select to be superintendent of schools if he (she or they) had the proper experience and training.
6. Whom you like best.
7. Who offered the best solutions to the problems discussed.

[3] The reader can recognize a similarity between this technique and the pooled judgement method for differentially weighting traits of a composite criterion described in Burtt, H. E. *Principles of employment psychology.* New York: Harper, 1942, 354.

8. Whom you would like to see as chairman or head of the department.
9. Who most motivated the others to participate in the discussion.
10. Who seemed most interested in the discussion.
11. Whose class you would like most to be in, if all the participants were teachers.
12. Whom you would select to address an audience of teachers.
13. Who should get the best grade for today's discussion.

For the "Spread N" rating procedure, the total number of votes received from all those who rated a participant were divided by the number of raters to obtain his mean rating. To obtain an individual's rating by means of the "3 Best minus 3 Worst" technique, the number of "worst" votes received from all the raters were subtracted from the number of "best" votes. For the ranking procedure, a participant's average rank was computed. As will be shown later, there seemed little value in computing mean ratings for each of the thirteen items, item by item, as they all seemed to be measuring the same variable—leadership status.

Agreement Among Raters

It was felt unprofitable to compute the 190 intercorrelations between raters because of the small number of ratees. Participants' total "Spread N" ratings in the first, second, fifth, and sixth discussions assigned by a given rater, were converted instead into ranks, and the average intercorrelation of all rank orders was computed to give a rough appraisal of the extent of agreement among judges (See Woodworth (5), p. 372 ff.). The average intercorrelations obtained for the first and last two discussions were .72, .61, .63, and .41 respectively.

The correlation between combined participant's ratings and combined observer's ratings of each participant was another measure of the extent of agreement among raters. When participants' ratings for the first 2

Table 1

The Rate-Rerate Reliabilities of Nineteen Judges Rating Participants in Discussions I and II and Six Weeks Later in Discussions V and VI

Judge	r	Judge	r	Judge	r
1	.66	8	.65	15	.63
2	.64	9	.53	16	.86
3	.90	10	.86	17	.84
4	.48	11	.69	18	.41
5	.79	12	.72	19	.53
6	.76	13	.76	20	.72
7	*	14	.67		

* Judge 7 failed to attend the last two discussions.

discussions were combined into 1 distribution, there was a correlation[4] of .90 between ratings assigned to participants, by participants, and by observers.

Retest and Rerate Reliability

Each judge's ratings of participants of the first 2 discussions were correlated with the ratings assigned by the judges to the same individuals when they acted as participants of the last 2 discussions 6 weeks later. The fifth discussion was among the same participants as the first, and the sixth discussion was among the same participants as the second. Table 1 lists the 19 rate-rerate reliability coefficients obtained. By converting these coefficients into Fisher's Z-function, a mean coefficient of .72 was obtained. There was a correlation of .87 between the total time participants talked in the first or second discussion, and the time they talked in the fifth or sixth discussion.

The 10 subjects of the 20 who participated in the first 2 discussions who received the highest ratings were placed in the third discussion. The other 10 subjects were placed in the fourth discussion. Despite some reversals, those in the highest leadership status in discussions I or II became the "leaders of leaders" in discussion III and the highest "followers" in discussions I or II became the "leaders of the followers" in discussion IV. Rank order correlations between participants' ratings in discussions I or II and III or IV were .78 and .76 respectively. The conclusion may be reached that despite the change in groups and restriction of range, leadership status tended to be generalized from one leaderless group discussion to another. There seemed to be consistency in both the behavior of participants and the ratings they received in the leaderless group discussion.

Interrelationships Among Variables

The number of votes assigned to participants of discussion I or II by all raters were combined, item by item, and correlated with the time participants spent talking. Table 2 shows the results obtained. When votes obtained on all 13 items were correlated with time spent talking, the coefficient obtained was .93. Because of the high correlations obtained between each item and time spent talking, it was thought unnecessary to compute the matrix of intercorrelations, as all seemed to be measuring the same factor, i.e. leadership status. The unidimensionality may have been due in part to halo effect.

Discussion V and VI showed similar results. The correlation between time and leadership status ratings was .86. Taking into account the high rate-rerate reliabilities mentioned previously, it seems that time

[4] Unless otherwise stated, correlations were obtained by the Pearson product-moment formula.

Table 2

Correlations Between Time Spent Talking in the First Two Discussions and Number of Votes Received on Each of Thirteen Leadership Items

Item	r	Item	r	Item	r
1	.91	6	.85	11	.86
2	.87	7	.92	12	.87
3	.90	8	.82	13	.92
4	.89	9	.90	All Items	
5	.85	10	.87	Combined	.93

spent talking and leadership status are closely associated and that this close association may be generalized from one leaderless group discussion to another.

Comparison Among Rating Methods

The effects of using a different rating technique were negligible. In discussion III there was an almost perfect linear relationship between "Spread N" ratings assigned by 20 raters, and the "3 Best minus 3 Worst" ratings. In discussion IV, an almost perfect curvilinear relationship of the order $y = a + b \log X$ was found to exist between ranking and the "Spread N" ratings, but whereas the former tended to distribute individuals evenly, the "Spread N" like the "3 Best minus 3 Worst" method tended to dichotomize participants. The "Spread N" seemed to scatter out the leaders more widely, while the "3 Best minus 3 Worst" best spread out the followers. From a realistic approach, on the basis of this extremely small sample, the "Spread N" would seem to be the most valuable as a technique for selecting from the top end of a distribution. Leaders were distributed by the "Spread N" in the same manner as on the variable of time spent talking. The correlation between leadership status and time spent talking was therefore highest with this rating method. Since the leaderless group discussion was designed primarily to discriminate among leaders, the "Spread N" best approaches the needs of the situation and the correlation obtained when using this rating method, does not appear to be spurious to any great extent.

Interpretation and Conclusions

Several hypotheses can be drawn for further investigation from the results obtained and from personal observations of the leaderless group discussion.

1. If a group is given a verbal problem, with suitable motivation to cooperate and achieve the goals relevant to the problem, a differentiation of function will occur within the group.

2. In a leaderless group discussion, one task may be assumed by several people; some tasks may be assumed by one; some tasks may not be performed at all. These tasks include initiation or formulation of the problems and goals, organization of the group's thinking, clarifying other individuals' responses, integrating responses of several individuals, questioning, motivating others to respond, accepting or rejecting other individuals' responses, outlining the discussion, summarizing, generalizing, obtaining the group's agreement and formulating conclusions.

3. Because of the verbal nature of the situation, the more tasks an individual assumes, the more time he is forced to spend in talking to the rest of the group.

4. It is assumed that those individuals who carry out the abovementioned tasks are perceived by others to be the leaders of the group discussion.

5. If the above hypotheses are correct, then the time an individual spends in talking in the leaderless group discussion is indicative of his status as a leader or follower in that group situation.

Received April 11, 1949.

References

1. Brody, W. Judging candidates by observing them in unsupervised group discussion. *Personnel J.*, 1947 **26**, 170–173.
2. Frazer, J. M. New-type selection boards in industry. *Occup. Psychol.*, 1947, **11**, 170–178.
3. OSS Assessment of Staff. *Assessment of men.* New York: Rinehart, 1948.
4. Taft, R. Use of the "group situation observation" method in the selection of trainee executives. *J. appl. Psychol.*, 1948, **32**, 587–594.
5. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

Performance on the File-Remmers Test, How Supervise? Before and After a Course in Psychology

Harry W. Karn

Carnegie Institute of Technology

A recently developed instrument for the measurement of the attitudes and understandings necessary for supervisory success is the File-Remmers questionnaire, *How Supervise?* The construction of the original forms of this test is described in an article by File (1). In a summary of the preliminary evidence on the validity of the device, File and Remmers (2) report significant increases in scores after supervisory training and significant differences in scores between successful supervisors and individuals by-passed because of lack of supervisory ability.

An indication that the test is lacking in universal validity is the study by Sartain (4). This investigator reports the test to have little or no predictive value for success in supervision in an aircraft factory. In an attempt to reconcile these findings with their positive evidence, File and Remmers (2) question whether Sartain's 40 supervisors in an expanded war industry are sufficiently typical to be used as cases for drawing general conclusions concerning the usefulness of tests in selecting supervisory personnel.

In view of the evidence reported to date, further studies appear to be in order before the question concerning the universal validity of the test can be settled. As a contribution toward this end, the following report is presented of an investigation designed to determine the effect of a psychology course upon college students' understanding of supervisory skills and practices as measured by the File-Remmers test.

Procedure

Forms A plus B of the test were administered under standard instructions to 108 students (104 males) in the College of Engineering and Science at Carnegie Institute of Technology during the first week of a first semester required course in psychology. About 98 per cent of this group consisted of students in their junior college year with the remainder consisting of irregulars in the sophomore and senior years. These students made up the *training group.* The same tests were administered to a comparable *control group* of 104 students (101 males) during the first week of a required course in English. During the last week of the semester both groups were again administered both forms of the test.

During the interim between testings the specific issues covered in the test were not discussed in either the psychology or English courses. Scores were not divulged nor the purpose of the investigation mentioned until after the final testings.

The training group consisted of five sections taught by four different instructors. This group took a three-credit course in general psychology. All of the instructors used standard psychological text books although the emphasis throughout the course was not upon text book content per se but upon the application of psychological principles to adjustment problems, particularly those likely to be encountered in human relations situations in industry. To this end, realistic case problems were dealt with and the student encouraged to solve them through the use of psychology and a systematic, analytical problem solving procedure.

The control group was made up of five sections of a three-credit course in English literature taught by five different instructors.

A feature of the investigation is the use of a control group with which to compare changes in test performance by the training or experimental group. This type of experimental design is mandatory if valid conclusions concerning the effectiveness of any training program are to be drawn. Changes in performance on the part of the experimental group can be attributed to the training program only if these changes are significantly different from any changes that may appear in the control group.

Results

Table 1 summarizes the essential statistical data for a comparison of scores made by the training and control groups on Forms A plus B of the test during the original and terminal testings. These data indicate a slight difference in mean raw scores between groups on the first test, an

Table 1

Comparison of Initial and Terminal Scores on Forms A plus B for Training and Control Groups

Group	N		Mean	S.D.	Critical Ratio*	r between Original and Terminal Scores
Training	108	Before	95.3	15.0		
		After	104.5	13.3	7.46	.61
Control	104	Before	97.1	16.9		
		After	98.9	17.9	1.70	.81

* Critical ratio of difference between groups on first test, .81. Critical ratio of difference between groups on terminal test, 2.54.

insignificant increase on the part of the control group on the retest and a highly significant increase of nine points of raw score on the retest by the training group. The nine point increase in mean raw score at this level is equivalent to a shift from about the 50th percentile to about the 70th percentile, according to the File-Remmer norms for Higher Level Supervisors (3). A comparison between retest scores of the training and control groups shows a difference of about five and a half points in mean score which is significant at nearly the one per cent level of confidence. This comparison, however, obscures the total gain made by the training group since this group had a lower mean (95.3) than the control group (97.1) at the time of the initial testing. The full extent of the gain made by the training group is evident in the ensuing comparative analysis which treats the data in terms of absolute differences in scores between testings.

A summary of the analysis made in terms of the differences between scores on the initial and terminal testings for training and control groups is presented in Table 2. These data reveal the average difference between the initial scores and the higher terminal scores for the training group to be significantly greater than a comparable measure for the control group.

Table 2

Comparison of Training and Control Groups in Terms of Differences Between Initial and Terminal Testings

	Training	Control
N	108	104
Mean of Differences	9.3 (gain)	1.6 (gain)
S.D.	12.8	10.9
Critical Ratio	4.73	

Because of greater reliability, both forms of the test are recommended by its designers over the single form. It is reported, however, that the single form yields scores which are sufficiently reliable for gaining information about a group as a whole. It appeared worthwhile, therefore, to compare the double form analysis from the present study with the data from single forms. Scores on Form B from the initial testings were compared with scores made on the same form on the terminal testings. Form B was used because any possible practice effects from having taken another form of the test would be equated on both initial and terminal testings.

Table 3 is a summary of the data from the single form analysis. There is little difference between mean scores of the two groups on the initial test, an insignificant increase by the control group on the retest, and a

Table 3

Comparison of Initial and Terminal Scores on Form B for Both Groups

Group	N		Mean	S.D.	Critical Ratio*	r between Original and Terminal Scores
Training	108	Before	50.3	8.0		
		After	54.1	7.5	6.61	.61
Control	104	Before	50.8	10.3		
		After	51.5	10.6	.85	.68

* Critical ratio of difference between groups on first test, .43. Critical ratio of difference between groups on terminal test, 2.04.

highly significant increase of four points of raw score on the retest by the training group. The increase of four points at this level is equivalent to a shift from the 55th percentile to the 70th percentile, according to the File-Remmers norms (3). The difference of nearly three points in mean score between groups on the retest is significant at about the five per cent level of confidence.

The full extent of the gains made by the training group on the single test form is revealed in the summary of the data presented in Table 4. These data show the average difference between the initial and terminal scores for the training group to be greater than the comparable measure for the control group at about a one per cent significance level.

Table 4

Single Form B Comparison of Training and Control Groups in Terms of Differences Between Initial and Terminal Testings

	Training	Control
N	108	104
Mean of Differences	4.1 (gain)	1.4 (gain)
S.D.	7.5	7.5
Critical Ratio	2.56	

In general, the results of the single form analysis corroborate those from the analysis of data from both forms although the differences for the latter attain a higher level of statistical significance than those for the former.

Discussion

On the assumption that a psychology course designed to improve understanding of the principles of human behavior in industrial situations actually accomplishes this goal, the superior terminal questionnaire per-

formance of the students having taken this course would indicate that the instrument is measuring those skills having to do with an understanding of the principles of the successful management of human relationships In support of this conclusion is the absence of significant improvement in scores on the part of a comparable group of students tested at the same times but without having taken the psychology course between testings. Since good supervision presumably requires a knowledge of human relations skills, the present study can be interpreted as evidence for the validity of the File-Remmers test as a means of measuring one aspect of supervisory success.

The present investigation is a contribution towards the establishment of the universal validity of the test since it deals for the first time with college students under academic instruction. Previous studies, which have shown positive gains with training, have been concerned with the effects of specific industrial training programs among on-the-job supervisory personnel. The demonstration of improved scores under a variety of training conditions indicates that the responses are the result of the application of basic principles to the problems rather than the acquisition of specific answers to the questionnaire items.

High scores on a test of the type under discussion are no guarantee that the individuals making such scores will be good supervisors. There must be additional evidence from other sources that such individuals will put into practice the knowledge they possess. The argument for the use of the test is still good, however, for obviously individuals cannot put into practice knowledge that they do not have. The final answer to the question of measuring and predicting success in supervisory situations will probably take the form of a composite index based on test scores, biographical data and on-the-job ratings by qualified observers.

Summary

A group of 108 college students in their junior year were administered Forms A plus B of the File-Remmers questionnaire, *How Supervise?* before and after a course in psychology. The same tests were administered to a comparable control group of 104 students before and after a course in English literature.

An analysis of the data in terms of mean scores for both groups on initial and terminal testings and in terms of differences between scores on the two testings shows significant gains in favor of the psychology group. This is true for both the double form data, i.e., the comparison of scores on Forms A plus B on initial and terminal tests and the single form comparison of scores on Form B only.

Received March 7, 1949.

References

1. File, Q. W. The measurement of supervisory quality in industry. *J. appl. Psychol.*, 1945, **29**, 323–337.
2. File, Q. W., and Remmers, H. H. Studies in supervisory evaluation. *J. appl. Psychol.*, 1946, **30**, 421–425.
3. File, Q. W., and Remmers, H. H. *How Supervise?* (revised manual) New York: 1948, Psychological Corporation, pp. 8.
4. Sartain, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *J. appl. Psychol.*, 1946, **30**, 328–339.

The Prediction of Accidents of Taxicab Drivers

Edwin E. Ghiselli and Clarence W. Brown

University of California

The long history of investigations concerned with the effectiveness of tests in the prediction of accident proneness among vehicle operators might lead one to suspect that a wealth of information exists on this subject. Examination of typical reviews of the literature, however, points up the fact that empirical evaluations are by no means extensive and are quite restricted with respect to types of tests, being almost wholly concerned with apparatus tests (**2**, **5**, **7**). There appears to be general agreement that several kinds of reaction time tests, particularly those involving more complex functions, are the most useful. The value of tests of sensory acuity and perception is less certain. In the paper and pencil test field only intelligence tests have had more than a cursory examination. The validity of this type of predictor is rather low, with a validity coefficient of the order of about .15.

In many situations apparatus tests cannot be employed in the selection of vehicle operators. Both initial and maintenance costs often are too high. In certain cases applicants must be tested in large groups, and in others personnel capable of operating apparatus tests are not available. Indeed, in some instances sheer lack of space for a permanent set up of testing equipment is an obstacle. However one might view the desirability of apparatus tests as compared with the paper and pencil variety, financial, administrative, and physical limitations may be the deciding factor in the choice of the type of tests that can be utilized. This leads, then, to a need for a closer scrutiny of the data available concerning the effectiveness of paper and pencil tests.

The low validity of intelligence tests cited earlier can be considered to assume some significance if only the very few superior individuals in a large number of applicants are to be considered. The Personnel Research Section of the Adjutant General's Office have reported eight validity coefficients for intelligence tests and nine for mechanical principles tests relative to performance on various types of road tests (**6**). In the large majority of these studies the coefficients were based on more than one hundred cases. For intelligence tests the validity coefficients range from .03 to .33, with a median of .18, and for mechanical principles tests the range of validity coefficients is from .00 to .40, with a median of .20.

Since driving skill might be expected to be related to safety of operation these coefficients at least are suggestive that paper and pencil tests might be helpful in predicting accidents. The present authors, together with E. W. Minium, investigated the usefulness for street car motormen of a variety of tests, including some of the paper and pencil variety (4). The criterion in this study was accidents for an eight-month period. Paper and pencil tapping and dotting tests were found to be the most valid, surpassing even apparatus tests (sensory acuity, distance perception, and simple reaction time). A coefficient of the order of about .35 is descriptive of the tapping and dotting tests. Tests of mechanical principles, judgment of distance by perspective, and judgment of linear distances were found to be less useful. For them a validity coefficient of .15 is representative.

Data such as the foregoing, meager though they may be, at least are encouraging. Further information regarding the effectiveness of paper and pencil tests would be most helpful in indicating fruitful lines for further research. The results of a single investigation should not be considered definitive but certainly would be a helpful addition to knowledge in such a relatively unexplored area. It was with this intent that the present investigation of the effectiveness of various types of predictors of accidents of taxicab drivers was undertaken.

Description of Predictors

Eight different speed tests and an interest inventory which yielded four different scores were used. The first five of these tests were the paper and pencil tests employed by Ghiselli, Brown, and Minium with street car motormen (4). Following is a description of each predictor.

Dotting. This test consists of a series of circles one-eighth inch in diameter, connected by lines, and irregularly spaced. The subject is instructed to place one dot in each circle. No pretest practice is given and the time limit is one-half minute.

Tapping. In this test the subject is presented with a series of circles of one-half inch in diameter and is instructed to put three dots in each circle. Again, no pretest practice is permitted and the time limit is one-half minute.

Judgment of Distance. The intent of this test is to measure capacity to judge distances between objects utilizing as cues only perspective and interposition. Each item is a schematic representation of a table top on which are placed four equally sized cubes. The positions of the cubes and the angle of view are different in the different items. The task is to judge which of three cubes is nearest to a designated key cube. The time allowed for this test is eight minutes.

Distance Discrimination. Each item in this test consists of a square in which there are three test points and a reference point. The task is to decide which of the test points is closest to the reference point. Various distracting lines wind among the points. The time limit is three and one-half minutes.

Mechanical Principles. This test consists of a series of pictorially presented problems illustrative of various mechanical principles. Most of the items are

concerned with the movement of vehicles and the operation of levers. Only simple mechanical principles are involved. The time for this test is eight minutes.

Numerical Problems. This is an arithmetic test involving the making of change and the computation of fares when various rates and lengths of trip are given. Five and one-half minutes are allowed for this test.

Speed of Reactions I. This test is an attempt to put in paper and pencil form a complex reaction test such as the Viteles Motorman Test. Each item consists of a square in which different letters appear in various spatial arrangements. Depending on the specific letters given in an item and their spatial arrangement, the subject makes a mark in one of five spatially differentiated circles placed below the square. This test is preceded by detailed instructions concerning the rules together with some examples for practice. On each page of the test proper the rules are given so the subject has them immediately available for reference. The time for this tests is four and one-half minutes.

Speed of Reaction II. This test is the same as Speed of Reactions I except that the subject can never refer to the rules once the test is begun. Four minutes are allowed.

Interest. The interest inventory has no time limit. It yields four separate scores and a total score which is simply the sum of the part scores. For each of the four scales there are 24 items. In each item the subject chooses between two different occupations. He is instructed to choose on the basis of interest and ignore such matters as pay, vacations, and the like. The first scale has to do with *Occupational Level*, and compares a job roughly at the semiskilled level with a higher level job. The correct answer is the lower level job. The second scale is concerned with *Outside Occupations*, the correct choice being a job which is done out of doors rather than indoors. The third scale attempts to measure interest in occupations which require dealing with the public, and is termed *Dealing with People*. The last scale, *Related Occupations*, compares jobs involving the operation of vehicles with other types of jobs.

In addition to these tests certain personal data concerning the drivers were available. This information consisted of age, years of formal education, years of previous experience driving taxicabs, and years of experience operating other types of vehicles, either commercially or in the armed services. Since frequently considerable reliance is put on these variables in hiring drivers they, too, were studied in relation to accidents.

Subjects

The subjects used in this investigation were 67 men who applied for work and were employed as drivers by a taxicab company during the same three-month period. All men took the tests prior to being hired, and to some extent their scores were taken into account in the decision regarding their employment. In the selection process a profile of test and inventory scores was plotted for each person and those individuals who showed marked deficiencies were rejected. Greatest emphasis was given to the numerical problems and speed of reactions tests and to the interest inventory scores. Approximately one out of five applicants were rejected on this basis. With the exception of the speed of reactions test all men took all tests. For the speed of reactions tests the number of cases is 57.

Criterion

Accidents are a notoriously unreliable index of human behavior (5). Furthermore, accident proneness is by no means a unique trait. Examination of the relationships among different types of accidents incurred by operators of public conveyances indicates that there is considerable specificity (1). The problem of setting up a criterion to measure safety of performance, then, is fraught with many difficulties. In the present investigation the situation was complicated by the fact that only the safety records for the first five weeks of employment were available. However, this period is particularly critical since it is during this time that supervisors' judgments concerning the drivers' skill are crystallized. Of the 67 drivers, 48 incurred no accidents during the first five weeks of employment, 17 were involved in one accident, and two men were involved in two accidents. With such a distribution of accidents it was impossible to determine validity relative to number of accidents. As a consequence the men were divided into two groups, the accident free men (48 cases) and the accident group (19 cases), and biserial coefficients of correlation were utilized as indices of validity.

Results

In Table 1 are given the validity coefficients for the various predictors using accidents of the drivers during their first five weeks of employment as the criterion. It is apparent from this table that with the exception of the dotting and tapping tests, and possibly the interest inventory, none of the validity coefficients can be considered particularly significant.

The validity coefficients of the first five tests, dotting, tapping, judgment of distance, distance discrimination, and mechanical principles, are approximately of the order found earlier for motormen (4). The pre-

Table 1

Validity Coefficients for Various Predictors in Relation to Accidents

Validity Coefficient	Predictor	Validity Coefficient	Predictor
.35	Dotting	.23	Occupational Level (Interest)
.47	Tapping	.23	Outside Occupations (Interest)
.18	Judgment of Distance	.11	Dealing with People (Interest)
.20	Distance Discrimination	.20	Related Occupations (Interest)
.11	Mechanical Principles	−.08	Age
−.09	Numerical Problems	−.10	Years of Formal Education
−.04	Speed of Reactions I	.00	Years of Experience Driving Taxicabs
.00	Speed of Reactions II		
.28	Total Interest Inventory	.07	Years of Commercial and Service Driving

dictive power of the first two tests is fairly substantial while that of the latter three is low. On the basis of the motorman study weights were developed for these five tests for computation of a battery score. The effective weights are as follows: dotting, and tapping each four, judgment of distance and distance discrimination each two, and mechanical principles one. Applying these effective weights to the scores of the 67 taxicab drivers the validity coefficient of the battery was found to be .69. None of the accident group fell in the upper 25% of scores. Undoubtedly this coefficient is fortuitously high and would not be obtained with another similar sample. For the motormen the validity of this battery was of the order of .35. A battery of this type certainly seems worthy of further study for the selection of operators of vehicles. It is apparent that the tapping and dotting items are the most important components of the battery. Remembering that the total testing time for these two tests is only one minute the validity coefficients are surprisingly high. They are, in fact, too high to be readily acceptable just on the basis of two investigations, and further evaluation of these tests is indicated.

The fact that scores on the numerical problems test were found to be unrelated to accidents is not unexpected. There is no reason to suppose that ability to solve arithmetical problems is related to capacity to drive a motor vehicle safely. However, the complete lack of validity on the part of the speed of reactions tests certainly was not anticipated. At least, superficially, these tests seem to measure very nearly the same abilities as those measured by complex reaction tests, such as the Viteles Motorman Test, which have considerable predictive power. Perhaps these findings may be taken to indicate that certain kinds of abilities measured by apparatus tests and important in safe performance cannot be measured by paper and pencil prototypes.

The interest inventory that was utilized was designed principally to indicate which applicants would tend to stay on the job as taxicab drivers and which would tend to leave. In the area where this study was made labor turnover in the taxicab industry was quite high. An analysis of the situation indicated that a large proportion of the persons who left employment did so because they found that they did not like the nature of the work or at least they preferred other types of jobs. Nevertheless, the validity coefficient of .28 is suggestive of the value of interest measures in the prediction of accidents. The particular inventory utilized in this investigation undoubtedly can be greatly improved and the fact that three of the scales (occupational level, outside occupations, and related occupations) yielded validity coefficients of .20 cannot be ignored.

In the hiring of workers considerable emphasis is given to the factors of age, education, and work history. The subjects in the present investigation differed markedly in these variables. The range in each

variable is as follows: for age, 21 to 53 years, for education, seventh grade to college graduate, for previous taxicab experience none to 6 years, and for all types of experience driving vehicles other than private cars none to 23 years. In spite of the wide range of individual differences none of these factors was found to be appreciably related to safety of operation. In view of the success with which personal data has been used in the selection of employees in other occupations the relationships here may well be subject to some doubt. It is quite possible that the relationships are curvilinear and thus would not be adequately measured by biserial coefficients. The numbers of cases in the study was considered too small to warrant any detailed study of curvilinear relationships. However, inspection of the tabulations suggested an optimal age of 30 years and an optimal educational level of 10 grades for safe performance on the job. No similar optimums could be noted for previous driving experience.

Discussion

Taken together with the findings of previous investigations, the results of this study indicate that accidents can be predicted by paper and pencil tests. No brief can be made for the position that such tests invariably will be effective, but at least it can be said that they can be as effective as apparatus tests. It is likely, of course, that a combination of paper and pencil and apparatus tests would give better results than either the one or the other type alone. However, it may be questioned whether for purposes of employee selection the inclusion of apparatus tests increases the goodness of prediction sufficient to warrant the increased costs. It is true that apparatus tests have considerable face validity and thus are meaningful to applicants. Nevertheless, paper and pencil tests of the sort used in the present investigation, while novel to the large bulk of applicants, were readily accepted by them. By means of proper instructions and judicious choice of items paper and pencil tests, too, can be made with satisfactory face validity.

The battery consisting of the dotting, tapping, judgment of distance, distance discrimination, and mechanical principles tests seems most promising as a selective device for the selection of safe vehicle operators. On two different groups of operators it has given acceptable results. One difficulty with the battery is that most weight is assigned to two very short tests. It is almost inconceivable that the dotting and tapping tests with their half-minute times could have reliability coefficients higher than .60, and, indeed, a lower figure would seem more reasonable. It is possible, of course, that they do measure abilities of particular importance to safe operation, but certainly more evidence than two studies covering only a total of some 220 cases is necessary before any great confidence can be placed in them.

Another difficulty with the battery utilized here, and in fact, with any paper and pencil tests, is their relative uselessness for training purposes. Scores on visual and reaction time tests are most helpful aids in safety training programs. Having estimates of an individual's visual acuity glare sensitivity, reaction time, etc., intelligent recommendations can be made and pertinent training initiated to compensate for any deficiencies. With motor vehicle operators Fletcher has been able to produce significant and long term improvements in driving safety by utilizing such tests and interpreting the significance of the scores to the individuals concerned (3). Probably the best that paper and pencil tests can do for training is to indicate which individuals should be given special attention in any safety training programs.

Summary

The scores earned by 67 taxicab drivers on eight paper and pencil tests and an interest inventory, together with certain personal data items, were studied in relation to safety of operation. Accidents during the first five weeks of employment formed the criterion in the computation of validity coefficients. Dotting and tapping tests were found to have the highest validity. Tests involving judgment of distances and of knowledge of simple mechanical principles yielded low validity coefficients. A combination of the scores on the five foregoing tests, weighted on the basis of evidence collected in another investigation with an entirely different group, yielded a battery which was found to have a validity coefficient of .59 for accidents. An arithmetic test and a paper and pencil test of complex reactions were found to be useless in predicting accidents. Interest measures showed some promise, particularly for scales of occupational level, outside occupations and related occupations. No significant relationships were found between the accident criterion and age, education, and previous driving experience.

Received April 1, 1949.

References

1. Brown, C. W., and Ghiselli, E. E. Accident proneness among streetcar motormen and motor coach operators. *J. appl. Psychol.*, 1948, **32**, 20–23.
2. DeSilva, H. R. *Why we have automobile accidents.* Wiley, 1942.
3. Fletcher, E. D. Capacity of special tests to measure driving ability. Mimeographed, undated.
4. Ghiselli, E. E., Brown, C. W., and Minium, E. W. The use of test scores for the prediction of accidents of street car motormen. Report to the Municipal Railway System of San Francisco, 1946.
5. Ghiselli, E. E., and Brown, C. W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948.
6. Personnel Research Section, Adjutant General's Office. Statistical Manual. 1944.
7. Viteles, M. S. *Industrial psychology.* New York: Norton, 1932.

Some Precautions in the Use of the Per Cent Method of Job Evaluation

William D. Turner

University of Pennsylvania

Section 5 of the computation for establishing factor comparison scales by the per cent method of job evaluation (**3**) involves a table of $F\%/J\%$ ratios (or estimates of relative job totals, see **4**, p. 156F) the columns of which have different totals. It may be assumed that such differences between column totals follow principally from fortuitous variation of intra- and inter-job factor patterns, since judgment errors contributing to such differences will tend to be cancelled out in each column total. The individual ratios in these columns are quite fallible estimates of relative job totals, because each ratio is based ultimately on two independent and fallible per cent ratings. An improved estimate of job totals could be had if there existed several such estimates for each given job, which could then be averaged with the expectation of cancelling out material portions of the judgment errors inhering in each. One's first impulse is to average the ratios in each given *row* of Section 5. But the aforementioned differences between the totals of the *columns* of these ratios indicate that each column is of a different order of magnitude, and that an estimate in one column is therefore not strictly comparable with one in the same row but another column. Accordingly, in Section 5 of the computation, four of the columns are multiplied by "Reduction Constants" so that their respective totals become equal to that of the remaining (M) column whose total is arbitrarily taken as a base. The results of such multiplications appear in Section 8 of the computation. The *rows* in Section 8 are then summed (which amounts to averaging, since each row contains the same number of ratios) to yield the column of "(8) Totals" appended to Section 4. These "(8) Totals" represent the improved estimates of job totals sought above.

Likewise, improved estimates of factor totals are obtained by applying "Reduction Constants" to the *rows* of Section 6, ($J\%/F\%$ ratios), so that the resulting *columns* in Section 7 contain relative factor totals of comparable orders of magnitude, which are in turn summed to obtain improved estimates of such totals. See the "Total" row in Section 7, whose contents are subsequently "converted" to adjust their general level of magnitude to that of the "(8) Totals" discussed above, before being appended to Section 3.

Consequences of Eliminating the "Reduction" Process

Hay (**2**) essentially proposes to eliminate the "Reduction" operation in question, and, hence, to average incomparable relative figures. If one had valid reason to weight differently, say, the columns in Section 5, one could quite properly depart from the writer's procedure. But there is no known rational basis for such differential weighting, and Hay's procedure accepts an irrational and adventitious weighting for which there is no theoretical defense.

Whether Hay's method can be justified in practice depends then on the magnitude by which his results depart from those yielded by the writer's method, and on the relative value of committee time lost in correcting the larger, ultimately detectable errors introduced by Hay's method.

When Hay's abbreviated procedure is applied to the data in Section 5 of the writer's article (**3**), and when the general level of magnitude of the resulting relative job and factor totals is adjusted to that of such totals in Sections 3 and 4 of the method, factor ratings by $F\%$ and $J\%$ comparable with those in Section 11 may be computed. When this is done, and when such totals and ratings are compared with those obtained by the complete computation, 45% of the 137 percentage differences between the results by the two methods are greater than zero. 28% of these percentage differences exceed 3; 21% exceed 4; 20% exceed 5; and 13% exceed 6. Three each of these differences exceed 7 and 8; and one each exceeds 9, 10, 11, and 12.

The significance of the foregoing percentage differences depends particularly on the frequency and magnitude of the larger ones relative to a job rating committee's judgment error. So far as the writer is aware, no one has determined the exact value of such an error for a given committee. However, the committee whose results illustrate the writer's article showed an average 5% percentage difference between its original and its reviewed ratings on about 300 jobs, when the review in question followed soon after a leveling off of the committee's skill. Percentage differences between the results of a subsequent review and the one mentioned can safely be estimated to lie between 2% and 4%. Assuming a 3% error as a compromise figure, about a third of the ratings which Hay's procedure would have supplied to this committee would have departed from the ratings produced by the original computation by amounts equalling or exceeding the committee's own emerging average judgment error, and several of them would have differed from the latter ratings by as much as three or four times this error.

Again, the foregoing discussion obviously assumes that ratings produced by the original computation approximate true rating values more

closely than do those produced by Hay's computation, and that the size and frequency of discrepancies between the ratings produced by the two methods signify the degree of fault in ratings produced by Hay's method. This assumption is made because, as noted above, Hay's computation involves the operation of averaging relative figures of different orders of magnitude. Such an operation augments the inevitable errors (of judgment) already present in the ratings produced by the original computational procedure. The rational "Reduction" process in Sections 5 and 6 of the original procedure minimizes this difficulty.

Statisticians will question why an *average* percentage discrepancy of 1.8% between the results by the two computational methods is important when the committee's *average* judgment error itself is probably as great as 3%. Since a job evaluation committee is not so concerned with achieving a small *average* error characteristic of its ratings in the aggregate as it is with minimizing its error in each particular rating it makes, those large errors actually introduced by Hay's method, and which the committee can come ultimately to detect, become important. There are 18 such discrepancies in the illustrative case, which are at least twice as great as the committee's average judgment error, and 4 of which are at least three times as great as this error. Only a few such errors in the original scales are all that are needed to throw many subsequent ratings out of line before the committee can become able to recognize their incorrectness, and the committee would then need considerable time to make the necessary corrections. Such committee time would be considerably more expensive than the hour or less gained by the computer using Hay's abbreviated procedure; and unless the discrepancies in question are ironed out, errors and borderline cases in job grading will be increased.

On the basis of the foregoing findings and conclusions, the writer sees fit to warn against the use of Hay's abbreviated procedure, and to recommend that the complete computation as originally set forth be followed in all cases.

Mathematical Relations Underlying the Per Cent Method

It should be observed that Hay's proofs that row totals correspond with job totals, and that reciprocals of column totals correspond with factor totals, obscure their underlying assumption that relative values of differing orders of magnitude can be validly averaged. The writer's present findings, and the marked differences between some of the "Reduction Constants" which appear in Sections 5 and 6 of the writer's article (3), emphasize the practical untenability of such an assumption and of the proofs which Hay bases on it.

In the writer's second article (**4**) the attempt was made to show some of the meaning of the per cent method's complete computations, without recourse to algebraic notation. An algebraic version of pages 155 and 156 of this article follows:

Let

r = any one of a number of factor ratings obtained during the use of established factor rating scales;

T_J = the total of factor ratings (r's) for any job;

T_F = the total of factor ratings (r's) for any one factor for any given group of jobs for which T_J's are also available;

$J = r/T_J$ = a "J value" (see Table II, p. 155, Reference 4) corresponding to a given r; and

$F = r/T_F$ = an "F value" (see Table III, p. 155, Reference 4) corresponding to the same r.

Then, $F/J = (r/T_F)/(r/T_J) = (r/T_F)\,(T_J/r) = T_J/T_F$.

But for any given factor grouping of r's, T_F would be constant, and may be regarded as equal to unity. Therefore, the *relative job totals*, or $T'_J = F/J$.

By corresponding proof, the *relative factor totals*, or $T'_F = J/F$.

An algebraic account of the more complex situation that arises when fallible $F\%$ and $J\%$ values are substituted for mathematically infallible F and J values, would be essentially descriptive. Since the algebraic expressions involved would be even less easily followed by most readers than is the verbal account in the writer's second article, the writer has not published such an algebraic description.

Weber's Law and Job Evaluation

Hay implies (**1**, **2**) that Weber's Law applies to factor comparison job evaluation. In brief, Weber's Law says that a discriminable difference between two physical stimuli bears a constant ratio to the level of magnitude of the stimuli themselves. The absence of any physical or objective measure of job values makes an application of Weber's Law to job evaluation data logically impossible. Hay refers to unpublished evidence of his own, to the effect that discriminable differences expressed in rating scale units bear constant ratios to the magnitude of ratings in factor comparison job evaluation, but apparently fails to consider the subjective nature of the scale units in which such differences *and* their corresponding ratings are necessarily expressed. Hence, the law which can properly express Hay's observations is not Weber's psychophysical Law, but a psychological Law of Per Cent Judgment which may be

written, $L/J = K$, in which

L = any discriminable difference (limen) expressed in per cent rating scale units or their equivalent; [1]
J = the point level of the scale rating characterized by L; and
K = an empirical constant.

It is evident that this Law of Per Cent Judgment involves no physical or objective measures, and that any similarity which it bears to Weber's Law is essentially mathematical; it is a psychological and not a psychophysical law. This latter distinction, which may seem to be purely theoretical, has an important practical implication for job evaluation. An assumption that Weber's Law applies to job evaluation data implies quite incorrectly that job values can be ascertained by some physical or objective method of measurement. Such an assumption would lend unwonted support to one of job evaluation's recurrent delusions which holds that "somewhere there exists the 'real objective truth' about job values." The Law of Per Cent Judgment emphasizes the inescapably subjective nature of the process of job evaluation.

Hay's proposed "limen" of 15% is actually determined by the Law of Per Cent Judgment formula given above, with L equalling the scalar distance from the *committee's* corresponding final ratings which includes at least 75% of committee members' individual preliminary ratings; with J equalling the former ratings; and with the K value multiplied by 100 to yield a percentage figure. However, the members of one of the rating committees whose work the writer has directed manifested in the aggregate a corresponding "limen" ratio equal to 9%, with corresponding "limens" for individual members lying between 6% and 11%. In order that a committee's judgments may express more fully the accuracy of which the committee and its members are capable, the writer recommends again (cf. **3**) that the size of geometric steps on factor rating scales be commensurate with a committee's emerging judgment accuracy rather than with a prescribed standard "limen" of 15%.

Received March 31, 1949.

[1] The term, "their equivalent" is meant to signify factor rating scales derived by the per cent method, or such scales as derived by Benge's method. The latter equivalent is justified by the extremely close correspondence (apparently limited only by judgment error) between ratings by the per cent method and Benge's method. Calling the present law one of per cent judgment relates it directly to the more controllable process of (per cent) judgment rather than to the salaries of equitably paid key jobs from which Benge derives his scales. The generally linear correlation between either per cent method or Benge method ratings and non-negotiated rates of pay, and the much closer agreement between ratings by the two methods in question, indicate that pay rate setting in the absence of job evaluation rests on a less accurate form of per cent judgment.

References

1. Hay, Edward N. Characteristics of factor comparison job evaluation. *Personnel*, 1946, **22**, 370–375.
2. Hay, Edward N. Creating factor comparison key scales by the per cent method. *J. appl. Psychol.*, 1948, **32**, 456–464.
3. Turner, William D. The per cent method of job evaluation. *Personnel*, 1948, **24**, 476–492.
4. Turner, William D. The mathematical basis of the per cent method of job evaluation. *Personnel*, 1948, **25**, 154–160.

Predicting Subject Grades of Liberal Arts Freshmen with the Kuder Preference Record *

Dorothy Terry Hake and C. H. Ruedisili

University of Wisconsin

The solution to the problem of the prediction of achievement is still in a preliminary stage. Interests are generally conceded to be important in determining success in any field of endeavor, but they are only one of many factors, such as intelligence, attitudes, and personality traits. The present study deals with the relationships between college achievement in specific subject areas and interest test scores. We have investigated the value of the Kuder Preference Record in predicting the first-semester grades made by freshmen at the University of Wisconsin.

Since the Preference Record was developed to measure interests, a high correlation with specific achievements would not necessarily be expected, but a low positive correlation with general college achievement has been found in previous studies. Crosby (**2**) has found positive correlations in the high .60's between the Preference Record Scientific scale and chemistry and biology grades, and between the Computational scale and accounting grades. His subjects were students scoring above the 90th or below the 10th percentile in each of the Preference Record scales, and therefore the relationship between interests and grades is not as high as it appears at first glance. If the total distribution of interest scores had been used, the correlations would have been considerably lower (**6**). Thompson (**9**) reports some success in predicting dental school success by using the Preference Record in conjunction with the MacQuarrie Test for Mechanical Ability. Bolanovich and Goodman (**1**) used the Preference Record to select women students for training programs in electronic engineering during the war. They found significant differences between successful and unsuccessful students. The former showed high scores on the Computational and Scientific scales, and low scores on the Musical and Clerical scales. Yum (**10**) found differences among students enrolled in the Physical, Biological, and Social Sciences, and also found low positive correlations with grade-point average.

Strong (**8**) mentions that since college courses are largely elective, the student does not choose courses in which he has no interest. He points

* Based on a dissertation submitted as partial fulfillment of the requirements for the M.A. degree at the University of Wisconsin, January, 1945.

out that interests, then, would not be important in determining success in elective college courses. University of Wisconsin freshmen at the time of this study, however, were much restricted in their choice of courses. The subject fields used were all freshman courses, and the usual freshman program consisted of four out of the five areas: English, Foreign Language, History, Science, and Mathematics. The problem of free-selection, then, is largely eliminated from the study and, presumably, measured interest scores might have some part in determining success in these subjects.

Procedure

The Preference Record was given to all Letters and Science freshmen who entered the University of Wisconsin in the fifteen-weeks summer session and the fall semester of 1943. The first-semester grades in five subjects and the over-all grade-point averages were obtained. The subjects included were English (N = 579), Science (N = 528), History (N = 402), Foreign Language (N = 477), and Mathematics (N = 201). All freshmen who took the Kuder Preference Record in the entrance examinations and completed the first semester, including one or more of the five courses, were included in the study; altogether, 594 students met these requirements (**3**). Men and women students were combined in this group since there were no separate norms for the Kuder Preference Record at the time this study was made (**4**).

Results and Discussion

The means and standard deviations of the Preference Record scales and of grades were computed for each subject area and for the whole course load (Grade-Point Average). These figures are in terms of raw scores for the scales and are not equated from scale to scale. The grades are in terms of grade-point averages, with 1.00 representing a grade of C. Differences in mean interest scores between subject groups are frequently large. The Mathematics group seems to be especially deviant in interest scores. It is relatively high in Computational and Scientific interest (39.32 and 66.89 respectively) and low in Social Service (69.17). This suggests that students taking Mathematics may have had more specific interest in the subject than did students taking other subjects. This seems likely, since Mathematics is often avoided as a difficult subject. Interest scores for the other subject groups are fairly similar, however, bearing out the preliminary hypothesis that these students' interests do not materially affect their choice of courses. The highest grade-point average was obtained by the Foreign Language group (1.37) and the lowest (1.15) by the English students. The highest S.D. occurred in the History group (1.46) and the lowest (0.82) in the English group.

Table 1

Correlations between Preference Record Scores and Grades

Scale	English	Science	History	Foreign Language	Mathematics	Grade-Point Average
Mechanical	−.15	.04	−.13	−.18	−.02	−.07
Computational	−.10	.10	.03	.03	.10	.04
Scientific	−.04	.18	.04	−.02	.10	.07
Persuasive	.08	−.05	.11	.00	−.02	.09
Artistic	.02	−.03	−.03	.01	−.02	.11
Literary	.25	−.01	.13	.12	.02	.13
Musical	.03	−.06	−.01	.10	−.10	−.01
Social Service	−.03	−.14	−.14	−.09	−.06	−.12
Clerical	−.07	−.04	.04	.07	−.01	.00

Table 1 contains the correlations between Preference Record scores and subject grades as well as grade-point average. The first column consists of the correlations between the grades of students who took English and their scores on each of the nine scales. The other columns show the same relationships for the remaining subjects. There are some expected relationships apparent in this table. The highest positive correlations are between Literary interests and grades in English, History, and the general grade-point average. The correlation between Science and the Scientific scale is in the expected direction. The inverse correlations between Mechanical interests and grades in English, History, and Foreign Language also seem reasonable. The Artistic and Clerical scales have the lowest correlations with grades in general.

The correlations among the individual scales on the Preference Record for the group as a whole are shown in Table 2. These intercorrelations, on the whole, are either low positive or negative. Relatively high positive correlations, however, are found between the Scientific and Mechani-

Table 2

Intercorrelations among Kuder Scores

	Mechanical	Computational	Scientific	Persuasive	Artistic	Literary	Musical	Social Service
Computational	.25							
Scientific	.64	.34						
Persuasive	−.20	−.19	−.35					
Artistic	.04	−.28	−.12	−.11				
Literary	−.31	−.15	−.31	.28	−.08			
Musical	−.28	−.18	−.23	.07	.16	.10		
Social Service	−.27	−.24	−.19	.09	−.10	−.02	−.02	
Clerical	−.20	.40	−.29	.07	−.20	.04	−.04	−.08

cal scales, the Literary and Persuasive scales, and between the Clerical and Computational scales. The highest negative correlation is that between the Scientific and Persuasive scales, while other fairly large inverse correlations are those between Mechanical and Literary, Scientific and Literary, Musical and Mechanical, Computational and Artistic, Scientific and Clerical. These same tendencies also can be noted in the intracorrelations found in five other groups, as described in the *Revised Kuder Preference Record Manual* (**5**).

The results of the Wherry-Doolittle Method of Test Selection included the progressive shrunken multiple-correlation coefficients, the uncorrected multiple-correlations, the name of the first scale which was not included in the battery because it caused a decrease in the shrunken multiple-correlations, and K (the coefficient of alienation). The highest of these shrunken multiple-correlations is only .3093 for General Grade-Point Average. One might expect to find that general school achievement can be predicted more accurately from interests than can grades in any one subject. Since the multiple-correlations for English (.2936) and History (.3022) are almost identical with that for the Grade-Point Average, however, there must be other factors involved. Possibly the relatively large size and heterogeneity of these groups may account for the higher multiple-correlations. The lowest multiple-correlation (.1997) is that using Mathematics as a criterion, and this group has the smallest N. It must also be remembered that this group seems to have relatively similar interests, and thus is a more homogeneous group. For the whole group, the test adding more chance error than validity was the Clerical scale; for the History, English, and Foreign Language groups it was Artistic; for Science it was the Persuasive scale, and for Mathematics the Computational scale.

Discussion

In general, it seems likely that interests, as measured by the Kuder Preference Record, are a relatively minor factor in predicting college achievement. Used alone, the Preference Record would probably be of little help. The addition of the Preference Record to college entrance examination batteries may be advisable, however, since interest measures may very well contribute significantly to the multiple-correlation obtained with the traditional aptitude and achievement tests. Further research which combines the Kuder Preference Record and other interest measures with aptitude and achievement scores is advisable before interest tests are rejected as being useless in predicting college achievement.

Summary

Scores on the Kuder Preference Record were compared with the grades obtained in five subject-fields by 594 students. The students

were first-semester freshmen in the College of Letters and Science at the University of Wisconsin. The correlations between grades and grade-point averages and scores on the Preference Record scales were computed and from these the shrunken multiple-correlations were obtained for each subject group and for the grade-point averages of the whole group by means of the Wherry-Doolittle Method of Test Selection.

1. The means and standard deviations of the raw Preference Record scores, with one exception, did not differ widely with respect to subject groups, indicating that interests are not an important factor in choosing freshman courses, and thus presumably are important in determining success.

2. The correlations between the scales and the grades and grade-point average are, on the whole, fairly low, but certain logical relationships between scores on the Preference Record and the subject groups can be noted. The Literary scale was found to have the highest positive correlations with the subject grades and grade-point average, while the Mechanical and Social Service scales showed fairly high inverse correlations.

3. The intercorrelations among the scales were low and negative, on the whole, but fairly high positive intercorrelations were found among the Mechanical, Computational, and Scientific scales, between the Literary and Persuasive, and between the Clerical and Computational scales. These same tendencies have been noted in other studies.

4. The results of the Wherry-Doolittle method show that a few of the scales, such as Mechanical, Scientific, Literary, and Social Service, were more useful than others in contributing to the multiple-correlations. The resulting shrunken multiple-correlations were all low. The largest obtained was that for the total grade-point average (.3093), and this approximated the best subject-fields, History (.3022), and English (.2936). It is concluded that interests, as measured by the Kuder Preference Record, may play a minor role in determining school achievement. In conjunction with other tests (achievement, scholastic aptitude, intelligence, attitude, etc.) the scores of this test may prove useful in the prediction of college grades.

Received March 15, 1949.

References

1. Bolanovich, D. J., and Goodman, C. H. A study of the Kuder Preference Record. *Educ. psychol. Measmt.*, 1944, **4**: 315–326.
2. Crosby, R. C. Scholastic achievement and measured interests. *J. applied Psychol.*, 1943, **27**: 101–104.
3. Froehlich, G. J. The prediction of Academic Success at the University of Wisconsin, 1909–1941. *Bulletin of the University of Wisconsin. Bureau of Guidance and Records of the University of Wisconsin, October* 1941.

4. *Intermediate manual for the Kuder Preference Record,* Chicago: Science Research Associates, 1944.
5. Kuder, G. F. *The revised manual for the Kuder Preference Record,* Chicago: Science Research Associates, 1946.
6. Peters, C. C., and Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: *McGraw-Hill Book Co.,* Inc., 1940.
7. Stead, W. H., and Shartle, C. L. *Occupational counseling techniques.* New York: American Book Co., 1940.
8. Strong, E. K. Vocational interests of men and women. *Stanford University Press,* 1943.
9. Thompson, C. E. Personality and interest factors in dental school success. *Educ. psychol. Measmt.,* 1944, **4**: 299–306.
10. Yum, K. S. Student preference in divisional studies and their preferential activities. *J. Psychol.* 1942, **13**: 193–200.

MMPI Personality Patterns for Various Occupations

E. E. Daniels and W. A. Hunter

VA Regional Office, Phoenix, Arizona

Two closely connected provocative questions regarding the role of dynamic personality patterns[1] in relation to vocational selection and personnel placement prompted the present investigation.

First, are there rather definite personality patterns which tend to gravitate toward certain of the multitude of occupations in the vocational world? Second, are there rather fixed "personality demands" in the various occupations which make up the work of man?

Industrial and employment interviewers, guidance personnel, teachers, etc., are aware of the numerous expressed reasons why individuals have gone to work in a particular job or embarked on a certain career. Menninger points up the problem as follows: "It would be interesting to examine how it came about that some people must do continuously what seems chiefly drudgery, while other people are able to do what seems to be pleasurable and even delightful work, if indeed it can be called work at all" (1). Is there a particular optimal pattern of personality factors for each occupation which when met contributes to occupational success and satisfaction? Will a standardized personality test be subtle enough to ferret out these patterns for research studies and even for practical application in job counseling?

A review of literature revealed a study made by Harmon and Wiener (2) who asserted, in using the MMPI, that personality characteristics appear to be of crucial importance in the actual choice of a vocation, a contention which appears to be a distinct aid in the prognosis of the success in training. In another study Verniaud (3) administered the MMPI to clerical workers, department store saleswomen and optical factory workers and concluded that saleswomen tend to make responses designated as "masculine," industrial women show definite trends toward hypomania and psychasthenia, while the clerical workers approach more nearly those responses which had been termed "normal." The present study used the MMPI (4) as an exploratory tool in an attempt to determine whether a relationship exists between the total personality "work needs" and the "personality demands" of occupations.

[1] The term "pattern" is used in the dynamic sense and substituted for the term "deviate" in the MMPI.

Methodology

The dynamics of this study were worked on over a period of 32 months in the Veterans Administration Regional Office, Phoenix, Arizona, and the raw data were collected from four VA Guidance Centers over the State. The conclusions are based on material drawn from a study of 893 veterans under both Public Laws 16 and 346. All cases had availed themselves of complete advisement and guidance, as set forth in the VA Manual of Advisement and Guidance (**5**), which culminated in the veteran's choice of an occupation. The MMPI categories were coded along with veteran's name on an IBM card, and the occupations which covered 97 category groupings from the Dictionary of Occupational Titles (**6**) were then obtained by IBM selection in terms of DOT code number from the master file of status cards. All cumulations were made on the IBM.

The group represented males of an average age of 23 years, constituting several racial groups from all sections of the United States and 90 per cent were high school graduates. For the purpose of coding on the IBM cards, the range of T scores on the MMPI was grouped at the center of a 10-point spread, i.e., at 55 for the range of scores 50 to 60.

After averages had been determined for each occupation and for each scale on the MMPI, tables were made up for each personality dimension separately. An F-test was made for the various MMPI characteristics. The F scores for the Mf, Pd, Sc, and Ma respectively were 5.85, 5.24, 2.79, and 3.36, which was very significant for Mf, Pd, and Ma, and significant for Sc, which means that the chances are less than 1 to 99 that so large an F could have occurred in a really homogeneous population. All semi-skilled and unskilled occupations, DOT codes 6–00 through 9–99, were omitted, leaving a total of 67 occupations. Using Fisher's Small Sample statistical technique, the significance of the obtained difference in the means between various pairs of occupations was calculated. In those cases where the difference was found significant there are about five chances, and for very significant about 1 chance in 100 that this could have occurred by random sampling. Not all pairs where there is a significant or very significant difference between the means are represented in the tables; only four of the MMPI scales out of a total of nine scales were selected for presentation in this article.

Results

Table 1 is composed of a list of selected occupations in terms of the average T score on the Masculine-Feminine, Psychopathic, Schizophrenic

Table 1

Mean T-scores on Four Scales of MMPI for Selected Occupational Groups

Occupational Group	N	Mf Scale	Pd Scale	Sc Scale	Ma Scale
Social Scientist	5	75.0	63.0	51.0	69.0
Physician	5	69.0	65.0	55.0	69.0
Author, Editor, Reporter	10	67.0	70.0	56.0	67.0
Chemist	7	65.0	56.0	52.0	57.0
Draftsman	13	58.8	57.0	55.8	57.0
Farmer, Livestock	5	47.0	57.0	55.0	55.0
Accountant	13	61.0	55.0	58.0	50.0
Medical Technician	8	55.0	63.0	48.0	56.0
Insurance Salesman	10	57.0	63.0	51.0	62.0
Athletic Coach	10	53.0	66.0	44.0	55.0
Auto Mechanic	33	52.2	62.0	52.0	61.0
Barber	4	65.0	55.0	52.0	57.0
Dancing Instructor	10	62.0	59.0	56.0	62.0
Farmer General	15	49.0	58.0	52.0	56.0
Commercial Artist	9	51.0	65.0	47.0	58.0
Social Worker	4	60.0	52.5	47.0	55.0
Clerk General	14	53.0	54.2	49.0	56.0
Auto Body Repairman	19	55.0	52.9	49.0	60.0
Managers	6	52.0	50.0	52.0	55.0
Lawyer	13	64.0	61.0	54.2	64.2
Typist	3	58.0	62.0	45.0	58.0
Radio Announcer	4	60.0	62.0	57.0	67.5
Electrical Repairman	7	52.0	58.0	52.0	49.3
Auto Upholsterer	6	57.0	63.0	55.0	53.3
Electrician	10	53.0	56.0	52.0	53.0
Kindergarten Teacher	5	57.0	63.0	55.0	63.0

and Manic scales. Tables 2, 3, 4, 5, and 6 giving complete results are omitted in this article because of cost.[2]

Attempts were made originally to group the occupations according to the major occupational groupings in the Dictionary of Occupational Titles. These broad groupings due to lack of homogeneity obscured the personality patterns of occupations within the group while showing no differences between groups. While occupational groupings may be made on the basis of one scale, it tends to obscure differences on other scales.

The data indicate that the means on the MMPI scales for the various occupations tend to scatter rather widely about the T score of 50; whereas, if the mean of all occupations combined is calculated, the mean approaches

[2] For Tables 2, 3, 4, 5, and 6 order Document 2694 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $0.50 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 × 8 inches) readable without optical aid.

rather closely the T score of 50 for each scale. The data would appear to indicate significant differences between the means of personality patterns as related to the various occupational objectives. Perhaps this would indicate that extensive differences in personality patterns exist between occupational groups at the extremes of the distribution of occupations obtained on the MMPI scales. It is found, for example, that the mean of personality scores for occupations taken from near the middle of the distribution is very significantly different from the means at either extreme. It is also noted that, since the MMPI scores for the various occupational groups tend to spread as indicated in Table 1, the statistically not significant differences indicate a degree of difference that possibly may be considered as placing each occupation near its optimal level on the scale for this particular personality pattern of the MMPI.

In contrast to a rather common interpretation of the MMPI that a score below 70 does not indicate a significant personality deviation, it is believed that any individual deviation from the mean T score, either positive or negative, on any personality scale is indicative of a certain tendency toward behavior in that direction, and that extremes, such as a critical T score of 70, are not necessary for the instrument to have definite meaning and application in the industrial field.

The *Masculine-Feminine* pattern on the Minnesota Multiphasic would appear, from these findings, to indicate a "work need" often requiring rechanneling in order that the occupational satisfaction of the basic Masculine-Feminine content of the total personality be achieved. This may be illustrated, for example, in the statistical difference between social scientist and farmer, livestock. The "work need" for a social scientist is an understanding of the problems of other human beings, the prerequisite of which is a high degree of sensitivity as seen in the Masculine-Feminine pattern. Another example of this need for redirection of the Masculine-Feminine component of the total personality can be seen in the occupation of physician, whose "work need" is expressed in his "bedside" manner, as compared to the low degree of this pattern in the occupation of draftsman, the difference between which is significant statistically. This dependency of occupational choice upon the Masculine-Feminine level as indicated by the MMPI has been empirically tested numerous times by us during vocational advisement and guidance by attempting to get a person who was interested in some occupation such as barber or beautician to consider the objectives of meat cutter or butcher, or vice versa. In all cases there has been a violent rejection of the consideration of the alternate objective. Statistical evidence and protocols seem to indicate that professions of a so-called highly cultural nature require as a fundamental "work need" a degree of Masculine-Feminine pattern approaching 70 T score on the Minnesota Multiphasic.

The *Psychopathic pattern* of the Minnesota Multiphasic would appear to be characterized by aggressiveness or by asocial behavior. Underneath this aggressiveness is the raw hostility and destructiveness as demonstrated by ample clinical evidence. This hostility originates from childhood reactions to authority in the family situation. Menninger points out that "The concept of work as drudgery which everyone experiences to some extent and which some persons experience to a very high degree, is bound up with this resistance to authority" (1).

The "work needs" of the individual personality with a high degree of Pd pattern may be illustrated in the choice of occupation such as author, editor, reporter, or athletic coach. "Purposeless destructiveness and aggressiveness may be molded and guided into the constructive activity of work" (1). This hostility may also be observed in those cases characterized by failure due to the unrecognized "work need" in the high Pd pattern of personality, as, for example, in the case of the veteran who was striving to achieve a father identification in his occupational selection by entering the same field as his father. The veteran was unsuccessful in his efforts until returning to the psychologist, wherein it became apparent from the evidence that hostility between the veteran and his father dated back to the earliest years. In attempting to alleviate the occupational blocking a counseling technique was applied whereby the veteran was urged to return East and spend his entire time, if possible, in close companionship with his father. This recommendation was accepted, and the veteran a number of weeks later returned with renewed interest and determination to succeed in his occupational efforts. The removal of the cause of competition with the father resulted in a freer expression of his occupational efforts in the same field. This would appear to illustrate the blockage of "work needs" which must be recognized if a rechanneling of childhood hostility into an occupational goal is to be successful.

Statistical findings in these data indicate that the difference between the occupation of athletic coach and the occupation of manager is significant, which might be interpreted to mean that the managerial occupations require a complete rechanneling of hostility in the direction of objectivity in management and administration on a more socially accepted level, whereas the athletic coach utilizes his aggressiveness mostly on a work level of physical effort (sports).

The *Schizophrenic pattern* on the Minnesota Multiphasic in this material appears to indicate a "work need" wherein the individual does not have to associate too closely with other people. The mechanism of isolation is well delineated in the clinical syndrome of Schizophrenia. The same mechanism appears to be effective to a lesser degree in influencing persons with a high Sc score in the choice of occupations, as exemplified in the difference between the draftsman as compared to athletic

coach, statistically a very significant difference. The occupation of draftsman may be considered as an isolating occupation, whereas the occupation of athletic coach required an essential capacity for dealing closely with other people. Another example which may be cited shows the difference in "work needs" for the occupation of typist, which is essentially a social and interpersonal occupation, as compared with the draftsman. Thus, occupations indicating a significant low degree of Sc pattern would appear to require a "work need" wherein the individual may satisfy his gregariousness, as compared to a high degree of Sc pattern wherein the occupation requires little association with others in the work situation.

The "work needs" as indicated in the *Manic pattern* of the Minnesota Multiphasic would appear essentially to be an outlet for enthusiasm and a high degree of overt activity. In the occupation of radio announcer, as compared to electrician, the difference is significant statistically, and may afford statisfaction in the occupation by providing an outlet for emotional and verbal expression. Again, this "work need" is clearly illustrated in the occupation of teacher, kindergarten, as compared with the occupation of electrical repairman, where the rechanneling of emotional content may be observed. Occupations which require dynamic behavior, such as that of lawyer, is another example, as contrasted with automobile uphoslterer, the difference between which is statistically significant.

Thus, many occupations seem to utilize and perhaps demand a personality pattern in which there is a great deal of spontaneity and enthusiasm expressed, whereas other occupations make little use of this personality pattern. Lewis' study on this problem asserted "that there is a relationship between occupational interests and personality tendencies" (**8**).

Discussion and Indicated Application

The "work needs" of the total personality have been presented in this investigation with the intent of stimulating further research. The dynamic relationship between the "work needs" of the total personality and the selection of occupation would appear to us to be significant. From these findings it seems desirable to scrutinize closely occupations in terms of "personality demands." *These "personality demands" once established could then be matched with the "work needs" of the total personality as indicated on the Multiphasic patterns in a manner similar to that already established in the occupational realm of job demands and physical capacities analysis.*

In this study the dynamics of the total personality are viewed in terms of their psychogenetic origins and their development through conditioned response, a learning phenomenon which may be easily observed in the mechanism of parental identification.

By utilizing this technique and viewpoint it seems to the authors that the Minnesota Multiphasic is a fairly sensitive instrument for measuring the total personality "work needs" in relation to the suitability of occupations having certain "personality demands."

Received March 18, 1949.

References

1. Menninger, K. *Love against hate.* New York: Harcourt, Brace and Company, 1942, pp. 136–137.
2. Harmon, L. R., and Wiener, D. N. Use of the Minnesota Multiphasic Personality Inventory in vocational advisement. *J. appl. Psychol.*, 1945, **29**, 132–141.
3. Verniaud, W. M. Occupational differences in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 604–613.
4. Hathaway, S. R., and McKinley, J. C. *Manual for the Multiphasic Personality Inventory*, New York: The Psychological Corporation, 1943.
5. Scott, I. D. *Manual of advisement and guidance.* Washington, D. C.: Veterans Administration, U. S. Government Printing Office, 1945, pp. 11–52, 83–180.
6. *Dictionary of occupational titles*, U. S. Department of Labor and U. S. Employment Service, Part I, Definitions of titles; Part II, Titles and codes; Part IV, Entry occupational classification, and supplement edition III, Washington, D.C.: U. S. Government Printing Office, 1939.
7. Lewis, J. A. Kuder preference record and MMPI scores for two occupational groups. *J. consult. Psychol.*, 1947, **11**, 194–201.

Correcting Special Ability Test Scores for General Ability

Abraham S. Levine

University of Minnesota

Most paper and pencil tests designed to measure special abilities or aptitudes correlate positively in varying degrees with tests of general ability or intelligence. This fact does not particularly detract and may actually enhance the predictive efficiency of special ability tests for occupations in which success is positively related to general ability. However, there are a large number of jobs particularly in the semi-skilled trades which do not require more than a modest level of general intelligence and for which a high degree of such ability has actually been shown to be related to high turnover rates. In these occupations the best predictors for the most part have been apparatus tests which are negligibly correlated with tests of general or verbal intelligence. For reasons of economy it is desirable wherever possible to administer group paper and pencil tests rather than individual apparatus tests. Therefore, if the effect of general ability could be partialed out, the utility of these contaminated paper and pencil tests as guidance and selection instruments for the relatively low I.Q. occupations may be increased.

The proposed method of correcting for the effect of general intelligence in a special ability test represents a simple application of the regression coefficient. A regression coefficient enables one to estimate the scores on a test if one knows the scores on another test with which it is correlated and the magnitude of this correlation coefficient for a given sample. .For the sake of illustration, let us choose two tests: (1) a general ability test such as the Tiffin and Lawshe Adaptability Test; and (2) a special ability test such as the Bennett Test of Mechanical Comprehension. Let us say that John Black obtained a score on the Adaptability test which was one and one-half standard deviations above the mean of a specified group, and there was a $+.50$ product moment r between the Adaptability and Mechanical Comprehension tests for this group. The best estimate of John's score on the Mechanical Comprehension test would be $.50 \times (+1.5)$ or a standard score of $+.75$.

The proposed correction for general ability substracts or adds to the special ability test score an amount defined by the size of the regression coefficient and the deviation of a general ability test score from the mean.

Thus, for example, if John Black actually obtained a standard score of .00 (mean score) on the Mechanical Comprehension test, the part contributed by his +1.5 standard score on the Adaptability test could be roughly corrected for by substracting .50 × (+1.5) from his Mechanical Comprehension standard score, thereby assigning him a corrected standard score on the latter test of −.75.

The above correction principle may be expressed by the following formula providing that all scores are converted into standard score units: Corrected Special Ability Score = Special Ability Score $-r \times$ General Ability Score.

The effect of using this correction formula is to raise the special ability score of an individual who is below the mean on the general ability test and to lower this score for an individual who is above the mean on the general ability test. The raising or lowering is of an amount proportional to the relationship between the two tests and the deviation from the mean on the general ability test. This correction formula serves to reduce the correlation between general ability test scores and corrected special ability test scores to zero, thereby eliminating the variance contributed by so-called general intelligence from a test designed to measure special ability or aptitude. Application of the correction would tend to obviate such disconcerting phenomena as bright but mechanically inept individuals scoring high on the Army Mechanical Ability Test and dull but mechanically gifted garage mechanics scoring low on this test by virtue of the aggravatingly high relationship between scores on the Mechanical Aptitude Test and the General Classification Test.

Incidentally, if corrected scores are to be computed for a large number of individuals, considerable economy may be effected by constructing tables which will enable one to read the corrected scores in either raw or standard score form directly from the table. For any given r a table can be easily constructed in which the special ability test scores are arranged in progression along the vertical and the general ability test scores along the horizontal, or vice versa, and the corrected scores found at the point of intersection in the table.

Corrected scores should be used only if the following conditions are fulfilled:

1. When the special ability test being considered is substantially correlated with a standard test of general intelligence for a particular sample. Otherwise, little is gained by introducing a correction factor which would necessarily be rather insignificant.

2. When success in an occupation for which the special ability test score is used as a predictor is not related to general intelligence. Otherwise, one would lose by eliminating the effect of a factor which is positively related to success.

3. Where there is an empirically demonstrated relationship between corrected scores and occupational success in excess of simply using the uncorrected special ability test scores. Since there is more work involved in obtaining a corrected score, it should justify itself by adding to the predictive efficiency for success in a given job. This is a crucial point since the rationale for the correction is primarily a practical one.

It is anticipated that corrected special ability test scores will find their greatest usefulness in the prediction of success in the semi-skilled trades and possibly in routine clerical jobs.

Received April 11, 1949.

The Rorschach Test in Industrial Selection

Audrey F. Rieger

Robert N. McMurry & Co., Chicago, Illinois

The place of the Rorschach inkblot test in clinical work has been well established, and some claim made that it is of value in vocational guidance. Another field to which the test can contribute valuable information is that of selection of industrial personnel.

The problem in selection is the choice of a worker who can fulfill the requirements of experience and ability for the job and who is a good risk for long-term employment. He must have the necessary skills and also be able to fit into the organization. Information about his adaptability to the job and to the company is very difficult to get, although some of it may be learned from interviews, references, and tests.

Personality questionnaires have often been used as aids in selection. The applicant, however, is frequently able to tell what the answers imply and finds it to his advantage to falsify his responses, if necessary, to give the impression he believes is favored for the position. The use of validating keys may permit detection of falsification, but they give little idea of the direction of the distortion.

A projective technique such as the Rorschach test makes falsification impossible, since the applicant can in no way determine what the examiner is looking for. The applicant must interpret the unstructured stimuli of the test in his own manner and is unable to determine how to produce a desired picture.

Projective techniques have some disadvantages, however. They are usually time-consuming and always require careful interpretation, a process which demands long and careful training. Therefore the cost of administering these techniques may make it impractical to use them, except for jobs which involve at least a moderate investment on the part of the employer.

As a result, few companies, with the exception of large organizations, can afford to add a Rorschach worker to their staffs. Providing them with access to the services of one on a consultant basis makes it possible for these employers to have the benefit of some information about the personality of the applicant when it is desirable and to pay for it only as it is needed.

For this reason, the services of a Rorschach worker were made available to the clients of a firm of personnel consultants. The test was

always given in conjunction with the regular selection technique, which is based on a Patterned Interview procedure.[1] Processing of the applicant included paper-and-pencil tests, the interview, and the Rorschach. The data derived from these sources were then weighted in order to make a recommendation to the employer with regard to the applicant's chances to be successful on the job.

In most instances the Rorschach test is of more value if given in advance of the interview; a brief report can then be made to the interviewer, with emphasis on clues which he can follow up in the interview. Occasionally, however, it was not possible to give the test prior to the interview. At such times, ratings for the job could be assigned independently from the interview findings and from the test results.

Table 1

Occupations of Subjects

Occupations	Frequency
Personnel Assistant	7
Personnel Director	4
Office Work	4
Sales	3
Engineer	2
Production Manager	2
Industrial Engineer	2
Merchandising Trainee	2
Market Research	2
Production Assistant	1
Reporter	1

Ratings based on the Rorschach results alone involved a comparison of the strengths and weaknesses reflected in the test results with the specific requirements of the job. For example, an applicant for an executive position was considered less promising if his record indicated difficulty in organizing abstract material or in controlling his impulses, whereas one who showed strength in these areas was more likely to be given a more favorable rating.

Under the special conditions of independent ratings, a total of thirty applicants were studied. Table 1 shows the occupations represented in the group.

Ratings for each of these subjects were made by the interviewer and by the Rorschach worker with the specific job in mind. Table 2 gives the distribution of the ratings, from 1 (superior) through 4 (reject).

[1] R. N. McMurry, *Handling personnel adjustment in industry.* New York: Harper, 1944, pp. 297 + xi.

The coefficient of correlation between the two sets of ratings is +.75 ± .05, a very significant result. A correlation of this magnitude indicates that use of either procedure will agree quite well with the results of the other. That the interview alone has great predictive value has previously been shown.[2] Using it as the criterion, it can be assumed that the results of the Rorschach are also valuable for industrial prediction.

It must be noted, however, that after only a brief period of using the test a correlation of this magnitude probably could not be achieved. The interviewer and the Rorschach worker had been associated over a relatively long period of time and were both familiar with the factors on which the recommendations were based and the methods used by the other in weighting these factors. This undoubtedly tended to raise the correlation.

Table 2

Scatter Table Showing Relation Between Rorschach Ratings and Interviewer's Ratings of Job Applicants

Rorschach Ratings	Interview Ratings			
	4	3	2	1
1	..	..	..	..
2	..	1	11	..
3	1	10	3	..
4	2	2	..	..

Nevertheless, it must be recognized that use of the Rorschach test by itself, a practice which is not recommended under any circumstances, would have led to selection of the best candidate in many more cases than chance alone would suggest. Such results would be found, however, only in instances where the Rorschach worker knew his instrument well.

The most efficient use of the technique is as a supplement to other employment procedures. It should at no time supplant them, as it cannot always assess properly the importance of various personality factors. In addition, it gives no information about the skills the individual possesses, his motivation to work, and several other factors which influence job success. Furthermore, unless care is used, the interpretation may be influenced by the bias of the examiner.

At the same time, the Rorschach offers unique help in learning many facts about job applicants (particularly at the higher occupational levels) and thus aids in improving prediction. The Rorschach test can be used objectively and offers much valuable information for selection.

Received April 10, 1949.

[2] R. N. McMurry, Validating the patterned interview. *Personnel*, 1947, 24, 263–272.

The Rorschach Test and Occupational Personalities

Audrey F. Rieger

Robert N. McMurry & Co., Chicago, Illinois

The question of differences in personality which may differentiate between occupational groups (and therefore aid in the selection of employees) has been raised, and some attempts have been made to answer it. Dodge, for example (**3**, **4**, **5**, **6**) found sales and clerical personnel had different patterns of scores on the Bernreuter. Paterson and Darley (**10**), using the same instrument, were unable to detect differences in their subjects. Verniaud (**14**), studying saleswomen, clerical workers, and optical workers, reports some differences in MMPI scores which she says correspond with differences in the occupational requirements.

Kaback (**8**), using the group Rorschach method, noted some statistically significant differences between pharmacists and accountants. Nevertheless, she concluded that neither group showed any generalized characteristics. A less exhaustive study using the same technique, that of Harrower and Cox (**7**), reports some differences between other occupational groups. Steiner (**13**) has summarized the Rorschach literature reporting studies of occupational groups.

The present investigation was designed to study personality patterns of certain specific occupational groups as reflected in the individual Rorschach test to determine if differences between such groups do occur and if the differences are meaningful in practical situations.

Subjects

The opportunity to make this investigation into occupational differences in personality arose in the course of routine procedures in the offices of an organization of personnel consultants. Applicants for positions with client organizations are interviewed (**9**) by one of the consultants and are given such paper-and-pencil tests as seem applicable to the positions for which they are being considered. In addition, the applicants are given an individual Rorschach test by the writer. The Rorschach was adopted for routine use in the employee evaluation program to give the interviewer a fairly objective portrait of the personality of the candidate and to aid in the evaluation of the information elicited in the interview, telephone checks with previous employers, and tests. On the basis of these data, the candidate is then rated with regard to his potential value as an employee.

As a rule, the candidates interviewed have previously been screened by the employer. The men doing this preliminary screening have been trained to be alert to clues indicating instability, inadequate intelligence, and other factors affecting success on the job. As a result, it seems likely that the candidates, most of whom would have been hired if expert advice had not been available, represent a better-than-average group of workers.

Hence these subjects cannot be considered as representative of applicants in their respective occupational fields. Moreover, the occupational classifications, based on the employers' job descriptions, may appear to be somewhat arbitrary, as some of the applicants lacked experience in the field. Since they were believed to have possibilities for the job, however, it seemed reasonable to include them as subjects for study. Some differences which might have occurred between more clear-cut occupational categories may have been obscured as a result, a fact which should not be overlooked in assessing the results of this study.

Table 1

Occupational Classification of Subjects

		Age		Years of Experience	
	N	Range	Mean	Range	Mean
Sales (technical)	55	19–48	29	0–20	3
Engineers	53	21–50	28	0–20	4
Supervisors, foremen	36	23–56	35	0–10	3
Administrators	64	24–48	35	0–18	5
Clerical workers	66	17–45	27	0–20	3
Personnel workers	24	22–45	32	0–11	2
Merchandising trainees	32	20–36	26	....	..
Miscellaneous	22	21–46	30	....	..

The jobs for which the applicants were being considered form the basis for classification into occupational groups. These jobs fall into six categories, with two additional miscellaneous groups. Table 1 summarizes some information about the groups. "Years of experience" noted in the table refers only to experience for the specific job, rather than years of work experience in general.

The supervisors were not being considered for eventual promotion into white collar jobs; they were men who had done well in the shop and were moving up. They had less formal education than the other subjects and did relatively less well on verbal tests, making their best scores on non-verbal items. The clerical group includes accountants, statisticians, and others doing similarly complex work. The last two groups,

the trainees and the miscellaneous, lack homogeneity as they include a wide range of occupations.

Methods

It was not possible to use as subjects only those who were recommended for employment, as the ratings were based on the results of the Rorschach test as well as on the information from the interview and test procedure and could not serve as criteria. Furthermore, the other data about the applicants were not uniform, the interviewing having been done by different individuals, the applicants having taken different tests, etc. As a result, none of the data except the Rorschach test scores could be used.

A large number of Rorschach scores were tabulated by occupational groups. Included were all those given weight in the orthodox interpretation of the results (1), such as color responses, the approach type, etc. In addition, the literature was reviewed for statements about occupational differences in personality which might be represented by Rorschach components; these were tallied. Finally, other scores which on an a priori basis may reflect differences between the groups were also studied.

Table 2 presents the means and standard deviations for all groups for the more important scores studied.[1]

In reviewing the results, the statistical reliability of the scores must be considered. Group differences may be minimized or completely obscured by lack of dependability of the measures. Unfortunately, no conclusive evidence has been put forth in the literature to answer this problem. In general, it is probably true that some of the Rorschach scores possess a high degree of reliability, and others are less dependable. It must also be recognized that ratios and difference scores, which represent relationships between imperfect measures, are less reliable than the component parts. Differences between groups might be hidden by such unreliability, and results with these scores must be taken with caution. Chief among these are FC−(C+CF), W%, etc.

Two statistical methods were used. The first involved testing the differences between the means of the groups for each score to determine if any of the differences were significant (CR at least 3). If such differences occurred consistently, it was planned to make up a composite picture of the worker in each field based on the means of the scores. Although this procedure is contrary to the basic idea of the interdependence of all behavior in the Rorschach test, it had had at least a limited success

[1] Table 2 may be ordered as Document 2652 from American Documentation Institute, 1719 N Street, N. W., Washington 6, D. C., remitting $0.50 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $0.50 for photocopies (6 × 8 inches) readable without optical aid.

elsewhere (11) and might aid in the development of occupational personality patterns to be adapted to selection and guidance.

The second method involved chi square tests of a number of scores which are more meaningful if interpreted in relation to each other, an approach difficult to manage in a study such as this where the clinical implications of many of the results must be overlooked for lack of information and an inability to deal with large numbers of subjects on an individual basis.

To equate one set of scores with another, a number of the measures were transmuted into a normalized scale of standard or T-scores, with identical means and standard deviations (2). A T-score on any one scale has the same relationship to the distribution of those scores in the total group as does the same T-score value on any other scale. The scores so treated were chiefly those making up the Approach Type (W, D, and Dd) and the Experience Balance (M and C).

Results

The attempt to find personality differences between the occupational groups had some significant results, but the Rorschach scores, as tested here, would fail to differentiate the groups in practice. Most of the differences appear to be related to variations in response total; for example, if R is high, W, D, or Dd is necessarily high in relation to the scores of other subjects with low R. Whether differences which are dependent on variations in R can be considered as real differences is a question which requires further study.

Tables 3, 4, and 5 summarize the significant differences between means.[2] The chi square tests support some of the results but fail to add much new information.

Only two groups of subjects tend to stand out from the remainder. These are the administrative group and the supervisors and foremen.

The administrators are characterized chiefly by their facility in producing and handling ideas (high R, low A%, etc.). Scores of this group indicate complexity of structure as well as lability and freedom of expression. Most of the areas in which these subjects differ from the remainder of the groups appear to be dependent on superior verbal facility, however, since the significance of the differences disappears when R is taken into account.

The supervisors and foremen form a fairly homogeneous classification and appear to be truly different from the other subjects. The group of supervisors, however, would probably stand out less noticeably if com-

[2] For brevity, the tables are omitted. The writer will be glad to supply the information upon request.

pared with similar workers; this is the only group of subjects not of the white collar or professional classes, a fact which must not be overlooked in reviewing their scores.

The supervisors are characterized by: limitation of ideas (low R), narrow range of interests (high A%), rigidity in judgment (high F + %), and restricted emotional life (low M and C). They are ill at ease in close associations with others (H-Hd low), and they tend to avoid contacts with others, even superficial relationships (H% low).

These restrictions may be explained by a number of factors. Probably most important is the fact that these men work chiefly with their hands and rarely deal with verbal concepts. Their weakness in verbal matters is evidenced by the relatively poor showing in verbal tests. In addition, the relatively impoverished background must be noted, with emphasis on the lower level of education. Finally there is the possibility that the personality may be reflected in the choice of occupation.

Although some traits seem to differentiate between the other groups (i.e., the salesmen seem to be more concerned with problems of health than are the other groups), these differences in scores may result from chance factors in the selection of the applicants rather than from specific group differences. The chief exception to this is the suggestion that the engineers show less interest in other people (H% low), a finding which corroborates the results of many other studies. In the case of the administrators also, the personality characteristics noted here can be related to some extent to the requirements of the job.

Discussion

The fact that the individual Rorschach test reflected variations between the occupational groups suggests that the Rorschach is sensitive to differences between groups. The test probably lacks reliability when used statistically to study differences between groups, however, and may therefore hide or distort some actual differences.

Another limitation is the selection of above-average rather than representative subjects. These may be more similar than would be subjects chosen at random.

Furthermore, the occupational classifications used in this study were not homogeneous categories of jobs. Although they were based on the employers' job titles, the duties under each heading varied widely. "Salesman," for example, might mean one who sold machinery or one who merely created good will for his employer's products. As a result, the heterogeneity of the jobs within each classification probably lessened the chances of turning up significant differences.

The fact that so few occupations could be differentiated in this in-

vestigation is of significance for selection and guidance. It is unlikely that the technique is wholly at fault in not turning up differences, since some consistent differences were noted. In some rare instances, such as cases where the individual possesses some special talent, the choice of occupation is determined by the talent. It is more likely that the occupation in which the worker spends most of his life is almost a matter of chance, determined by opportunity, rather than an end result directed by specific personality structure.

If this is true, no occupation can be said to draw people of similar personality makeup, although it may influence them to the extent that they later appear similar. There are a few exceptions to this, such as certain research fields (**12**). In general, however, any single personality pattern can be fitted into a number of jobs which may appear to differ greatly in demands on the individual; this is indicated by the great overlapping of scores between groups, even when the means differed significantly. No single personality type can be associated with any of the occupational groups, nor can it be assumed that any particular type of personality occurs to excess in any occupational group.

Recommendations for hiring must be based not on a general pattern for an occupation but on the specific requirements of the job and its place within a functioning organization. Here the Rorschach test can be of great value in pointing out the applicant's strengths and weaknesses, with due consideration to the part he will play in the particular organization and without concern for a generalized occupational pattern.

Summary

A study of several occupational groups by means of the individual Rorschach test showed a few statistically significant differences between groups. The only important result is the distinction found between those who deal with verbal concepts (chiefly administrators but including salesmen, engineers, clerical workers, and personnel workers) and those who work with their hands (supervisors and foremen). Personality patterns cannot be reliably used for placement, selection, and guidance. These findings (the lack of patterns) should not be construed as a denial of the importance of the descriptive elements of the Rorschach results in selection and guidance.

Received February 23, 1949.

References

1. Beck, S. J. *Rorschach's test.* New York: Grune & Stratton, 1944–1945, 2 vol.
2. Cronbach, L. J. A statistical method for treatment of limited patterns of scores. Unpublished MS, Univ. Chicago, 1948.
3. Dodge, A. F. Social dominance and sales personality. *J. appl. Psychol.*, 1938, **22**, 132–135.

4. Dodge, A. F. What are the personality traits of the successful sales person? *J. appl. Psychol.*, 1938, **22**, 229–238.
5. Dodge, A. F. What are the personality traits of the successful clerical worker? *J. appl. Psychol.*, 1940, **24**, 576–586.
6. Dodge, A. F. Characteristics of good clerks. *Person. J.*, 1942, **20**, 324–327.
7. Harrower, G. J., and Cox, K. J. The results obtained from a number of occupational groupings on the professional level with the Rorschach group method. *Bull. Can. Psychol. Assn.*, 1942, **2**, 31–33.
8. Kaback, Goldie Ruth. *Vocational personalities.* Teach. Coll. Contr. Educ., No. 924. New York: Columbia Univ. Press, 1946.
9. McMurry, R. N. *Handling personnel adjustment in industry.* New York: Harper, 1944.
10. Paterson, D. G., and Darley, J. *Men, women, and jobs.* Minneapolis: Univ. Minn. Press, 1936.
11. Rieger, Audrey. Rorschach analysis of adolescent groups. Unpublished MS, Univ. Chicago, 1945.
12. Roe, Anne. Rorschach study of a group of scientists and technicians. *J. consult. Psychol.*, 1946, **10**, 317–327.
13. Steiner, Matilda E. The use of the Rorschach method in industry. *Rorschach Res. Exch.*, 1947, **11**, 46–52.
14. Verniaud, Willie Maud. Occupational differences in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 604–613.

A Factorial Analysis of Arm-Hand Precision Tests *

Robert H. Seashore, Frank J. Dudek, and Wayne Holtzman

Northwestern University

Steadiness and precision of movement have long been thought to be a significant variable in various kinds of motor performance. Emphasis had been placed on the prediction of various skills by the use of various psychomotor tests and in the work done thus far there has been a tendency to accept a general factor or component of "steadiness." Seashore (3) sums up the conditions necessary for such a factor as follows: "According to the hypothesis of a group factor for steadiness, those coordinations which emphasize accuracy (precision or steadiness) while minimizing speed and strength should cluster together. Such steadiness tests, in wide variety, should intercorrelate moderately or highly, and show no correlations with speed and strength tests."

Various studies have presented evidence supporting the hypothesis of a group "steadiness" factor. Spaeth and Dunham (5) working with 73 army men, studied the relationship between Dunlap's test of precision in thrusting (1) and rifle target shooting. The correlation between the two tests for subjects ranging from poor to expert marksmen was .61, a very significant relationship. Seashore and Adams (4) found test intercorrelations of .45 or higher with a battery of five steadiness tests (postural sway with eyes closed, rifle muzzle sway when sighting, hand tremor, stylus thrusting at holes, and stylus held stationary in holes). Humphreys, Buxton, and Taylor (2) reported intercorrelations ranging from .37 to .69 with a median coefficient of between .52 and .55 between thrusting steadiness, stationary steadiness, an ataxiameter, and rifle sway. Relatively little research has been reported concerning the nature of the factors responsible for differences in performance on various tests measuring precision of movement and steadiness.

Purpose of Study

The present study was an attempt to determine the nature of factors underlying performance on seven measures of visuo-motor co-ordination.

* This study was carried out at Northwestern University as part of a larger project under the direction of Dr. R. H. Seashore. It was subsidized by the Office of Naval Research under its policy of encouraging basic research. The opinions and interpretations expressed, however, are those of the authors. The authors wish to acknowledge their indebtedness to Douglas Ellis, Richard Hetke, and Clarence Forsberg, who collected the experimental data for the second group of subjects.

These measures emphasized precision of movement of the preferred arm and hand. The tests used in this battery were selected with several considerations in mind: 1. they should be relatively uninfluenced by strength and speed; 2. they should be relatively free from the effects of muscular fatigue; 3. it should not be possible to get a high score by "trick" performances, and, 4. there should be little practice effect.

Tests Used

Seven tests were selected for inclusion in the battery. It is recognized that these seven tests do not sample, in all probability, the entire range of steadiness measures. However, it was not possible to include more variables in this battery because of time limitations. The more promising tests as determined from the analysis to be described will be included in another battery in an attempt to study and define more completely the domain of "steadiness." The seven tests are described below:

1 and 2. The Universal Ataxiameter: This test was designed to measure the horizontal and vertical components of involuntary movement of the hand and forearm. The subject attempted to hold a wooden tab or handle as motionless as possible. Movements were magnified by means of a leverage system and photo-electric cells recorded the amount of movement made. Horizontal and vertical components were scored separately. Five trials of 15 seconds each were administered in each cycle of tests.

3. Seashore Photoelectric Target Register (Revised): This test emphasized aiming and constant adjustment of a circular beam of light to a target. From a mirror which was on the end of a rod controlled by the subject the beam of light was reflected into a small hole. The beam of light activated a photo-electric cell which recorded the time the individual was "on the target." If the aim was perfect (i.e., the circle of light completely covering the hole) the counter recorded 10 counts per second—if the beam was only partly on the target the counts per second were correspondingly less. Five trials of 15 seconds each were administered in each cycle.

4. Straight Tracing Test: This test was a modification of the V-slot tracing test described by Whipple. At a controlled speed the subject drew a wire stylus between two brass plates without any base plate. The path formed by the brass plates was a converging one and the direction of the hand-movement was toward the body. If the stylus touched either side of the path it activated a very sensitive Potter Electronic Counter which counted at a rate of 60/sec. during time the stylus was in contact. Five trials were administered in each cycle.

5. Curved Tracing Test: This test was a variant of the straight trace. The path was of the same width throughout, but it was irregularly curved. The subject moved a wire stylus from left to right along this curved path. Scores were obtained on the Potter Electronic Counter as mentioned above. Ten trials of 15 seconds each were administered.

6. Sine-Curve Rod Tracing Test: In this test the subject moved a ring stylus of $\frac{7}{16}$ inch inside diameter along a brass rod of $\frac{3}{8}$ inch diameter. The rod was bent vertically in the form of a sine curve. The direction of arm-movement was from left to right. Time of contact between the ring and the rod was recorded by the Potter Electronic Counter. Speed of movement was controlled so that each trial required about 30 seconds. Five trials were administered per cycle.

7. Three-dimensional Rod Tracing Test: This test was constructed and scored similarly to the previous test, but the rod was bent into an irregular shape in three dimensions. The ring stylus had a somewhat larger inside diameter (1 and ⅛ inches) to permit greater freedom of movement. Five trials requiring approximately 40 seconds each were administered during each cycle.

8. Thrusting Steadiness Test: This test was a modification of Dunlap's test used by Seashore, Adams and others. The subject thrust a stylus into a hole in time with a metronome at a rate of one thrust every two seconds. Holes of three diameters were used, with S making 10 thrusts into each size hole in five different trials. The score was the number of thrusts made without contacting the side of the hole.

Method

This battery of tests was administered under a cycle plan. On one day the subject went through two successive cycles. Each cycle of tests required approximately 40 minutes to administer. Forty-eight hours later the subject returned and repeated two more cycles of the complete battery. In this way it was possible to obtain measures of reliability during any one testing period and for test-retest periods. Reliabilities are indicated in the last three columns of Table 1. The first two columns contain uncorrected, test-retest reliabilities for cycles combined in various ways. The last column is an estimate, corrected by the Spearman-Brown formula, of the reliability of the total score (i.e., all four cycles). The test-retest reliabilities range from .54 for the horizontal component of the ataxiameter to .85 for the three-dimensional rod test. These co-

Table 1

Intercorrelations, Means, S.D.s, and Reliabilities of Tests in Steadiness Battery (As computed for Group II, N = 100.)

		Product-moment *r*'s.									Reliability		
Variable	No.	1	2	3	4	5	6	7	M	S.D.	Cycles 1+2 vs 3+4	Cycles 1+3 vs 2+4	Of All Four Cycles*
Atax. (Hor.)	1								5.19	2.91	54	67	76
Atax. (Ver.)	2	47							4.43	2.68	75	84	89
Targ. Regis.	3	44	44						5.95	2.65	79	83	89
Str. Trace	4	39	20	40					3.10	2.76	79	82	89
Curv. Trace	5	30	04	26	61				4.77	3.47	82	95	95
Sine Trace	6	31	−06	26	57	72			4.43	3.12	77	89	91
3-dim. Trace	7	22	−07	34	37	72	83		4.08	3.68	85	82	91
Thrust	8	31	35	43	66	38	41	49	4.03	2.64	78	88	91

* Corrected according to the Spearman-Brown formula.

efficients indicate a rather high degree of stability for day-to-day measurement of precision of movement and steadiness.[1]

Results

The tests in this battery were administered to 100 volunteer subjects selected from elementary psychology classes at Northwestern University. All were right-handed males. Since stability of day-to-day measures seemed fairly high the scores used for analysis were the sums of scores on all performances for any given test. Product-moment correlations were then computed for these scores. These intercorrelations are presented in Table 1. The variables are, in general, positively correlated. No correlations are significantly negative though several are not significantly greater than zero.

Table 2

Centroid and Rotated Factor Loadings of Tests in Steadiness Battery

		Group II (N = 100)								Group I (N = 39)			
		Centroid Loadings					Rotated Loadings			Rotated Loadings			
Variable		I	II	III	h_c^2	h_c^2	I	II	III	I	II	III	h^2
Atax. (Hor.)	1	56	31	22	49	46	23	27	58				
										20*	10*	77*	64*
Atax. (Ver.)	2	37	58	24	54	53	−06	18	70				
Targ. Regis.	3	60	32	25	53	52	25	27	62	22	03	77	63
Str. Trace	4	78	09	−50	83	86	23	88	10	−03	66	11	45
Curv. Trace	5	72	−40	−09	68	68	68	46	−03	27	45	46	48
Sine Trace	6	74	−53	−03	83	84	81	41	−08	84	20	29	83
3-dim. Trace	7	72	−56	17	85	86	89	24	00	83	17	14	74
Thrust	8	69	17	−17	55	53	25	61	31	19	66	30	57

* In the data for Group 1 the horizontal and vertical scores for the Ataxiameter were combined into a single score.

The matrix of intercorrelations was subjected to a centroid factorial analysis. The results of this analysis are presented in Table 2. Most significant here are the two matrices of rotated factor loadings. The data for Group I were obtained in a preliminary investigation carried out by Holtzman in the summer of 1946. Only 39 individuals were used as subjects for this study. Nevertheless, the intercorrelations resulting from this sample were factorially analyzed with the resulting factors shown in

[1] Note that the coefficients of reliability which compare trials 1 and 2 of the first day with trials 3 and 4 of the second day are somewhat lower than the correlations between trials 1 and 3 vs. trials 2 and 4, which tend to balance out the diurnal vairation. G. Paulsen, *J. appl. Psychol.*, 1935, Vol. 19, pp. 166–79 and pp. 29–42 found, as we do, that there was a lower intercorrelation between trials on successive days than on the same day, on one test of hand steadiness. However there was still a very significant correlation between days.

the right-hand part of the table. In the present study data were gathered for 100 individuals (designated Group II). It is noteworthy that the factorial picture is so similar from group to group. Major factor loadings in each sample have been enclosed in boxes. It is apparent that the factor structure is practically identical from sample to sample.

Three factors were adequate to account for the correlations in each sample. When these were rotated with criteria of simple structure and positive manifold in mind the factors resulting from each analysis could be identified as similar. These factors were identified and named as follows:

Factor I has primary loadings on the three-dimensional rod test and on the sine-curve rod test. It has a major loading as well for the curved tracing test for Group II, but not for Group I. This factor seems obviously to be associated with steadiness and precision of movement which involves spatial components in two or more planes.

Factor II has major loadings in the straight tracing test and in the thrusting steadiness test. It will be recalled that in the straight tracing test the subject moved a stylus along a straight path toward his body. In the thrusting test the subject thrust a stylus toward a hole away from his body, but in a very similar path. This factor, it would seem, is associated with precision of movement in a restricted plane.

Factor III has major loadings for the target register and for the ataxiameter scores. The common component here is involuntary movement of the arm and hand. This factor seems to be what is usually thought of as "steadiness."

An examination of the factor loadings suggests that simple structure has not been ideally achieved and that the factors may be somewhat correlated. Correlations among the factors were estimated graphically by determining the cosine of the angular separation between oblique vectors representing them. Factors I and II are moderately correlated. Factor III does not seem to be highly correlated with the other two. It would seem, then, that stationary steadiness is relatively independent of steadiness where movement is involved. The correlation between factors I and II should not be surprising, since both involve movement. The number of dimensions within which movement takes place seems the important difference. Factor I included those tests calling for movement in a three-dimensional space, while Factor II was restricted more or less to movements in a restricted plane—or two-dimensional space. There is another possible interpretation based on the amount of movement involved. Those tests involved in Factor II require relatively shorter movements than do those appearing in Factor I. (The only anomaly here is the curved trace test which in this respect is more like the tests appearing on Factor II.) In this way it might be that Factor I represents tests in-

volving a general bodily orientation and control whereas Factor II tests represent finer adjustments within but one member of the body (the hand and arm). This hypothesis remains to be investigated.

Summary

The results of this investigation suggest several important considerations.

1. The intercorrelations among various tests all of which presumably depend on steadiness and precision of movement cannot be adequately accounted for by postulating a single factor of steadiness. There are, it would seem, various components influencing performance on these seemingly similar kinds of tasks.

2. There may be yet other factors accounting for scores made on steadiness tests measuring only involuntary kinds of arm and hand movement. This is suggested by the differences between the common factor variance and the reliabilities of the tests measuring involuntary movement in the present battery. While the reliabilities would indicate from 80 to 90% of the variance accounted for, common factors seem to account only for about 50% of the test variance. This would mean that about one-third of the variance is attributable to some specific factor or to a common factor yet to be identified by including these tests in another battery with other types of tests.

3. The results indicate that stationary steadiness (or involuntary movement of the arm and hand) is not highly related to precision of movement. Those factors, however, which involve spatial components are related to a greater degree.

These results and conclusions have implications in the area of selection in certain industrial areas. Many kinds of tasks are recognized as depending upon "precision," "steadiness," or "coordination." However, it would seem that there are several identifiable components—the isolation of which would improve our ability to select for the particular components which are represented in particular jobs. Thus—a lathe operator and drill-press operator might not require the same kinds of "steadiness."

Received March 14, 1949.

References

1. Dunlap, K. Improved forms of steadiness tester and tapping plate, *J. exp. Psychol.*, 1921, **4**, 430–433.
2. Humphreys, L. G., Buxton, C. E., and Taylor, H. R. Steadiness and rifle marksmanship, *J. appl. Psychol.*, 1936, **20**, 680–688.
3. Seashore, R. H. An experimental and theoretical analysis of fine motor skills, *Amer. J. Psychol.*, 1940, **53**, 86–98.
4. Seashore, R. H., and Adams, C. R. The measurement of steadiness: a new apparatus and results in marksmanship, *Science*, 1933, **78**, 285–287.
5. Spaeth, R. A., and Dunham, G. C. The correlation between motor control and rifle shooting, *Amer. J. Physiol.*, 1921, **56**, 249–256.

What Do Readership Studies Really Prove? *

H. P. Longstaff and G. P. Laybourn

University of Minnesota

Almost every publication seems to have "just made a reader study" on the basis of which it seems to be able to prove that it "leads the field in readership." Such violently conflicting claims have become so common that many a reader, advertiser, and publisher—confused by this contradiction of alleged facts—has begun to ask: "What do readership studies really prove?" One such perplexed publisher set out to try to answer this question.

Purpose

The purpose of the study conducted by the Putman Publishing Company was to call attention to fallacies inherent in "readership studies" as they have been commonly conducted and to suggest the need for more careful scrutiny of the results of such studies. It was *not* the purpose of this investigation to set up an entirely new, flawless technique for the study of relative readership. This was a purely analytical investigation of readership-study methods and techniques.

Procedure

To appraise the validity of the "orthodox" type of readership study, a procedure was devised to compare the relative readership standings of three industrial magazines on the basis of three different readership-study techniques which yielded, respectively, (1) the number of readers based upon the number of "mentions" [1] obtained in response to an original questionnaire employing "orthodox" readership-study techniques, (2) the number of readers corrected for "votes" [2] obtained in response to a follow-up questionnaire, and (3) the number of readers corrected for

* This paper is a condensation and revision of an investigation conducted by Mr. R. L. Putman of the Putman Publishing Company, and is published with their permission. The original study entitled "We Made a Reader Survey" was published by the above mentioned company. Copies of the original report are available (without charge to industrial advertisers and their advertising agencies) upon application to the Research Department, Putman Publishing Co., 737 North Michigan Avenue, Chicago 11, Illinois.

[1] *Mention* refers to the naming of a magazine in reply to the question, "What magazines do you read?"

[2] *Vote* refers to a "YES" response to the question, "Do you read this magazine?"

votes and "disqualifying negative comments" [3] obtained in response to the follow-up questionnaire.

This procedure entailed the use of three different questionnaires: (1) an original "orthodox" type of questionnaire, sent to 1,000 known readers of Magazine *A*, asking "What magazines do you read?"; (2) a follow-up questionnaire, sent to those who failed to mention Magazine *A* in response to the original questionnaire, asking "Do you read Magazine *A*?" and providing for comments on Magazine *A;* and (3) a follow-up questionnaire, sent to those who failed to mention Magazine *B* and/or Magazine *C* in response to the original questionnaire, asking either "Do you read Magazine *B*?" or "Do you read Magazine *C*?" and providing for comments on either Magazine *B* or Magazine *C*.

The First Questionnaire. The original questionnaire appeared on the letterhead of an independent manufacturer, who supposedly was attempting to determine in what industrial magazines he should place some advertising, and requested the addressee to write down the names of the industrial, business, or trade magazines which he reads. This first questionnaire was mailed to 1,000 "known readers" of Magazine *A* whose names and addresses had been copied from response slips[4] which readers[5] had taken from issues of Magazine *A*. Thus, Questionnaire No. 1 provided a check of what known readers say they read, after they had proved their readership of one publication, without knowing that the questioner knew anything about what they had read.

The Second Questionnaire. Since it was found that over one-half of the "known readers" of Magazine *A* who replied to Questionnaire No. 1 failed to mention Magazine *A*, it was decided to send them a follow-up questionnaire in order to discover why they had failed to do so. This second questionnaire appeared on the letterhead of the same manufacturer, who supposedly wondered whether the addressee's failure to mention Magazine *A* in replying to the first questionnaire was merely an oversight, and requested the addressee to indicate whether or not he read this magazine and provided a space for him to comment upon it. Thus, Questionnaire No. 2 was sent to those persons who, in replying to the original questionnaire, had failed to mention Magazine *A*.

[3] *Disqualifying negative comment* refers to a comment which indicates that a respondent who replied "YES" to the question, "Do you read this magazine?" actually does not read it.

[4] These slips had been inserted into copies of Magazine *A;* readers filled in subjects on which they wished more information, signed their names, and mailed the slips to the publisher.

[5] These readers were not necessarily subscribers to Magazine *A:* in several cases it was apparent that these inquirers had sent slips taken from copies of the magazine received by someone else.

The Third Questionnaire. Since it was found that nine out of every ten respondents replying to Questionnaire No. 2 reported that they read Magazine *A* (despite their failure to mention Magazine *A* in response to Questionnaire No. 1), it was decided to duplicate as closely as possible the conditions under which the second questionnaire had been sent out asking about Magazine *A* by sending out a third questionnaire asking about Magazine *B* and Magazine *C*. In order that no one would receive more than one follow-up questionnaire, Questionnaire No. 3 was not sent to those who had been sent Questionnaire No. 2, i.e., those who had failed to mention Magazine *A*. Thus, the third questionnaire was sent to those who, in replying to the original questionnaire, had failed to mention Magazine *B* and/or Magazine *C* but had mentioned Magazine *A*. Questionnaire No. 3 duplicated the conditions of Questionnaire No. 2 as closely as possible with the letter as nearly as identical as possible. Each questionnaire dealt with only one of the two magazines, Magazine *B* or Magazine *C*, and requested the addressee to indicate whether or not he read this magazine and provided a space for him to comment upon it.

Results and Discussion

The results of this investigation will be discussed in terms of (1) the response to each of the three questionnaires, (2) the effect of combining the results of the original questionnaire with the results of the two follow-up questionnaires, and (3) the interpretations necessitated by the comments accompanying the replies to the follow-up questionnaires.

The data obtained in response to each of the three questionnaires are presented in Table 1.

The First Questionnaire. In response to the 1,000 letters mailed to "known readers" of Magazine *A*, 585 replies were received,—a 58.5% response. Referring to Table 1, it will be noted that, of these 585 respondents "1,000% salted" for Magazine *A*, only 47.3% mentioned Magazine *A*. Thus, on the basis of the original questionnaire, Magazine *A* ranked third in readership.

The Second Questionnaire. Of the 291 who had failed to mention Magazine *A* in their replies to the first questionnaire and who were addressed with Questionnaire No. 2, 222 replied to the second questionnaire, —a 76.2% response. Referring to Table 1, it will be noticed that, of these 222 respondents, 91.4% replied "YES" to the question, "Do you read Magazine *A*?"

Why did so many of these respondents fail to mention Magazine *A* in replying to the first questionnaire? How could so many "change their minds" in replying to the second questionnaire? Surprisingly, 150 of those replying to Questionnaire No. 2 made some sort of comment—

Table 1

Response to Each of the Three Questionnaires

Original "Orthodox" Questionnaire: "What magazines do you read?"

Replies	1st Questionnaire					
	Magazine *A*		Magazine *B*		Magazine *C*	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
Mentioning as Read	277	47.3	309	52.8	295	50.4
Not Mentioning as Read	308	52.7	276	47.2	290	49.6
	585	100.0	585	100.0	585	100.0

Follow-up Questionnaires: "Do you read this magazine?"

Replies	2nd Questionnaire		3rd Questionnaire			
	Magazine *A*		Magazine *B*		Magazine *C*	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
YES	203	91.4	23	41.0	30	52.6
"Occasionally"	5	2.2	6	10.7	5	8.8
NO	13	5.9	27	48.3	21	36.8
No Reply	1	.5	0	0.0	1	1.8
	222	100.0	56	100.0	57	100.0

surprisingly, because such comment, while suggested, was not specifically asked for. This voluntary comment of these readers was perhaps the most revealing part of the response to the second questionnaire. Some typical comments follow:

> "I have received Magazine *A* for nine years and have asked for more information numberless times."
>
> "Overlooked this originally—I know of at least 15 other men in our organization who read Magazine *A* regularly."
>
> "Sorry to have overlooked Magazine *A* as this is really one of my favorite magazines along with Magazine *E* and really like it very much."

The Third Questionnaire. Of the 74 who had failed to mention Magazine *B* in their replies to the first questionnaire and who were addressed with Questionnaire No. 3, 56 replied to the third quetsionnaire,—a 75.6% response; and, of the 74 who had failed to mention Magazine *C* in their replies to the first questionnaire and who were addressed with Questionnaire No. 3, 57 replied to the third questionnaire,—a 77.0% response. Again referring to Table 1, it will be observed that, of those replying to Questionnaire No. 3, 41.0% of those who had failed to mention

Magazine *B* in the first questionnaire replied "YES" to the question, "Do you read Magazine *B*?"; and 52.6% of those who had failed to mention Magazine *C* in the first questionnaire replied "YES" to the question, "Do you read Magazine *C*?"

Effect of Combining Results

We have just seen that the percentages of "YES" replies to the question, "Do you read this magazine?" asked in the follow-up questionnaires, were 91.4, 41.0, and 52.6, respectively, for Magazine *A*, Magazine *B*, and Magazine *C*. Taking these percentages of each publication's "replies-failing-to-mention" in response to the original questionnaire, we note the following:

	Percentage Saying They Read in Response to Follow-up Questionnaire	Number Failing to Mention in Response to Original Questionnaire	Additional Readers
Magazine *A*	91.4	308	282
Magazine *B*	41.0	276	113
Magazine *C*	52.6	290	153

Then, adding the results of the follow-up questionnaires to the results of the original questionnaire, we arrive at the following:

	Number of Readers from Original Questionnaire	Number of Additional Readers from Follow-up Questionnaires	Total Readers
Magazine *A*	277	282	559
Magazine *B*	309	113	422
Magazine *C*	295	153	488

Thus, contrasting the results of the original "orthodox" study with the final figures obtained from all three questionnaires, we note the following inversion in the ranks of the three publications:

	Results from Original Orthodox Study		Figures from All Three Questionnaires	
	Readers	Rank	Readers	Rank
Magazine *A*	277	3rd	559	1st
Magazine *C*	295	2nd	448	2nd
Magazine *B*	309	1st	422	3rd

It should be noted that these final figures are *not* presented as accurate measurements of the relative readerships of these three magazines. Rather, it is believed that these final figures, contrasted with the figures of the first "orthodox" questionnaire, give evidence of the fallacies inherent in such "orthodox" readership studies.

Interpretations Necessitated by the Comments

What is perhaps the most revealing part of this entire investigation is found in the comments which were made by those replying to the follow-up questionnaires as to their readership of, and their opinions of, these publications.[6]

Comparison of Comments of Those Replying to Follow-up Questionnaires. The comments accompanying the replies in response to Questionnaire No. 2 (regarding Magazine *A*) and to Questionnaire No. 3 (regarding Magazine *B* or Magazine *C*) may be categorized as follows: (1) *Favorable*—i.e., the comment definitely reveals the respondent's approval of the publication and/or his actual readership of it; (2) *Negative*—i.e., the comment definitely shows that the respondent either does not actually read the publication regularly or does not feel that it is valuable to him; and (3) *Non-committal*—i.e., the comment tells nothing definite as to the respondent's actual readership or his personal opinion of the magazine's value to him.

Table 2

Comparison of Comments Accompanying Replies in Response to Follow-up Questionnaires

	In response to: 2nd Questionnaire		3rd Questionnaire			
	Magazine *A*		Magazine *B*		Magazine *C*	
Replies	No.	Per Cent	No.	Per Cent	No.	Per Cent
With Comments	150	67.5	27	48.2	44	77.1
Without Comments	72	32.5	29	51.8	13	22.9
Total Replies	222	100.0	56	100.0	57	100.0
Comments	No.	Per Cent	No.	Per Cent	No.	Per Cent
Favorable	102	68.0	6	22.2	10	22.7
Negative	11	7.3	18	66.6	29	65.9
Non-committal	37	24.7	3	11.2	5	11.4
Total Comments	150	100.0	27	100.0	44	100.0

Referring to Table 2, it will be seen that Magazine *A* elicited a strikingly larger percentage of "favorable" comments and a strikingly smaller percentage of "negative" comments than did either Mazagine *B* or Magazine *C*. From this tabulation there seems to be strong evidence, first, in the case of Magazine *A*, that actual comments show far greater

[6] A complete tabulation of all comments is presented in the original report, *We made a reader survey*, pp. 22–43.

active readership than mentions on the original questionnaire showed; second, in the cases of both Magazine *B* and Magazine *C*, that actual comments show far lower active readership than both mentions on the original questionnaire and votes on the follow-up questionnaire showed. Thus, it would seem that, in a readership study in which a publication's readership depended upon the number of mentions it receives, there would be a marked tendency for the readership of Magazine *A* to be underappraised and the readerships of Magazine *B* and Magazine *C* to be overappraised.

Do "Negative" Comments Disqualify "YES" Votes? An examination of the comments reveals the fact that some of those who voted "YES" on either Questionnaire No. 2 or Questionnaire No. 3 ("Do you read this magazine?") commented to the effect that they really do not read the publication in question. Typical of this type of comment are the following:

> "My recent reading, I am ashamed to say, has been sadly neglected because of a heavy work load. Consequently I have temporarily passed up this magazine. In my opinion it is a good magazine."
>
> "I read this magazine occasionally but have never inquired about items of interest in its advertisements. It takes too much time to read the magazine and carefully go through its advertisements."
>
> "Not too helpful in my line of work but have obtained some information from it at various times."

If we accept the principle that such negative comments disquality the "YES" votes, then, for each of the publications, we may deduct from the number of respondents voting "YES" when asked in the follow-up questionnaire "Do you read this magazine?" the number of such respondents whose comments disqualify their "YES" votes, yielding the number and percentage of "YES" votes corrected for "disqualifying negative comments." Following the method of calculation outlined above under "Effect of Combining Results," if we take these corrected percentages of each publication's "replies-failing-to-mention" in response to the original questionnaire, we may obtain the number of additional readers from the follow-up questionnnaires corrected for disqualifying negative comments. When these are added to the number of readers obtained in response to the original questionnaire, we arrive at the total number of readers obtained in response to the original questionnaire corrected for both votes and disqualifying negative comments in response to the follow-up questionnaires. These figures together with the ranks of each of the three magazines are presented in the last two columns of Table 3, which compares the relative standings of the three publications on the basis of the three methods of analysis employed.

Table 3

Comparison of the Relative Standings of Three Publications on the Basis of Three Different Readership-Study Techniques

	Relative Standings on Basis of: Number of Readers Based on "mentions" [1] Obtained in Response to Original Questionnaire Employing "orthodox" Readership-Study Techniques		Number of Readers Corrected for "votes" [2] Obtained in Response to Follow-up Questionnaire		Number of Readers Corrected for Votes and "disqualifying negative comments" [3] Obtained in Response to Follow-up Questionnaire	
Publication	Readers	Rank	Readers	Rank	Readers	Rank
Magazine *B*	309	1st	422	3rd	398	3rd
Magazine *C*	295	2nd	448	2nd	412	2nd
Magazine *A*	277	3rd	559	1st	550	1st

[1] *Mention* = the naming of a magazine in reply to the question, "What magazines do you read?"

[2] *Vote* = "YES" response to the question, "Do you read this magazine?"

[3] *Disqualifying negative comment* = comment which indicates that a respondent who replied "YES" to the question, "Do you read this magazine?" actually does not read it.

It should be noted that none of the figures contained in Table 3 are presented as accurate measurements of the relative readerships of these publications. These contrasting figures, however, do suggest that the further these studies are carried, the greater the discrepancy between the original "orthodox" technique results and the final figures.

What Do Readership Studies Really Prove?

The variations that are revealed in these figures would seem to lay down a challenge to the commonly accepted belief that one can measure readership by asking "What magazines do you read?" Both the "votes" and the comments obtained in the follow-up questionnaires of this investigation reveal the influences of the human tendencies to say we read what we feel we are expected to read, to boast of what we read beyond what we actually do read, and to protect ourselves from possible adverse criticism by stating that we keep up with what we think is accepted as "required reading." It seems apparent, therefore, that asking people "What do you read?" may measure the relative effectiveness of puslishers' promotion, publicity, and propaganda over many years, but does not necessarily measure readership.

Summary

1. In order to discover to what extent the results of "orthodox" readership studies are dependable, an attempt was made to compare the

relative readership standings of a number of publications as determined by three different readership-study techniques.

2. In the "orthodox" type of study employing a questionnaire which asked "What magazines do you read?" only 47.3% of those replying mentioned Magazine *A* even though practically everyone to whom the questionnaire had been sent was a "known reader" of Magazine *A*. On the basis of this original questionnaire, Magazine *A* ranked third in readership, Magazine *C* ranked second, and Magazine *B* ranked first.

3. When a follow-up questionnaire was sent to those who, in replying to the original questionnaire, had failed to mention Magazine *A*, Magazine *B*, or Magazine *C*, asking "Do you read this magazine (Magazine *A*, Magazine *B*, or Magazine *C*)?" the relative readership standings obtained in the original questionnaire were reversed.

4. When the comments made on the follow-up questionnaires were taken into account, the readerships of the three publications in question were changed still further.

5. It would seem apparent from this investigation, therefore, that the burden of proof rests on those who conduct "orthodox" readership studies to prove that their figures are measuring actual readership.

Received April 11, 1949.

Psychological Factors in Instrument Reading. II. The Accuracy of Pointer Position Interpolation as a Function of the Distance Between Scale Marks and Illumination *

Walter F. Grether

Aero Medical Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio

and

A. C. Williams, Jr.

University of Illinois

The reader of instruments is normally expected to obtain values of greater precision than the graduations placed upon the instrument scale. To accomplish this he must interpolate, that is, estimate the relative distance of the pointer from the two scale marks between which it falls and assign an appropriate value to this position. The accuracy with which this can be done obviously limits the precision with which any given scale can be read. The accuracy of such interpolation, moreover, will be influenced by several variables in the scale design and the conditions of reading. For a prediction of reading precision obtainable with different instrument designs under various conditions of viewing the effect of the significant variables must be known. In the present experiment the accuracy of pointer position interpolation was studied as a function of (a) diameter of the dial; (b) angular separation of the scale divisions; and (c) simulated day- versus night-viewing conditions. It will be shown in the presentation of the results that the first two variables can be reduced to a single one, namely, the length of the arc (in visual angle or inches) between scale marks.

A problem in scale design to which the present investigation is particularly relevant is concerned with the question of how finely a scale should be divided in order to provide maximum reading accuracy. In an investigation by Loucks (5) the legibility of tachometer dials was investi-

* This experiment was carried out at the University of Illinois by Dr. A. C. Williams, Jr., under a "dollar-a-year" contract with the USAF Air Materiel Command. Dr. W. F. Grether proposed the study, designed and procured the necessary dials, and prepared the present report. The basic data have been presented previously in Army Air Forces Aviation Psychology Program Research Report No. 19, Chapter 7, and in USAF Air Materiel Command Memorandum Report No. TSEAA-694-1.

gated using rather short exposure (0.75 sec.). For three dials with graduations of 100, 50, and 20 RPM respectively the percentage of reading errors increased as the value and size of the graduations decreased. From this finding it might be concluded that placing the graduations rather close together will decrease rather than increase reading accuracy. The findings of Kappauf, Smith, and Bray (**3**), and Kappauf and Smith (**4**), in experiments where the exposure interval was not limited, disagree with those of Loucks. In their experiments dials graduated in units gave greater reading accuracy and speed than dials of the same size but graduated in 5- or 10-unit steps. However, the superiority of 1-unit over 5-unit graduations was rather small and not at all in proportion to the increased number of graduation marks. These latter results are in agreement with those of an investigation by Grether (**1**) on the reading of clock dials. With one minute as the criterion of reading accuracy, dials with 1-minute graduations gave higher reading accuracy than similar dials with only 5-minute scale marks.

A possible explanation can be offered for the discrepancy between the findings of Loucks (**5**) and later investigators. It is quite probable that as the number of scale marks is increased more eye fixations are required to make each reading. By limiting the exposure time and consequently the number of eye fixations Loucks may have favored those dials with more widely spaced graduations.

In the study of scale designs it is helpful to distinguish between two general types of errors encountered in dial reading studies. There are first the precision errors or errors of interpolation. These can never exceed in magnitude the value of the smallest interval on the scale. The other type may be called comprehension of interpretation errors, in which an incorrect value is assigned to the graduation mark against which the pointer is being read. Comprehension errors are frequently very large and are usually some multiple of the minor, intermediate, or major scale divisions. In a study by Grether (**2**) of altimeter reading, for example, most of the errors were of this latter sort, with errors of 1000 feet being particularly common. It is important to recognize that many of the dial reading studies up to the present have been concerned only with the interpolation type of errors, when in actuality the larger comprehension errors are far more serious in practical instrument reading situations. It is quite possible that the presence of a large number of graduation marks on a dial may greatly increase the probability of large comprehension errors and thereby nullify the precision which a finely graduated scale makes possible.

The spacing of the divisions on a scale is usually not a variable concerning which the instrument designer has a free choice. Normally the

physical length of the scale, the range of values to be covered, and the desired accuracy of reading are fixed by the particular application. Based upon these requirements the designer must then select values (usually 1, 2, 5, or decimal multiples of these) for his scale increments which will give him reasonable spacing between graduation marks.

The aim of the present study was to provide the instrument designer with data from which to predict how the accuracy of readings will be affected by the physical length of the interval into which he sub-divides a scale. Measurements were made under two lighting conditions comparable to those under which aircraft instruments are viewed. Emphasis in this study was placed on interpolation errors. The more complex comprehension type of errors were recorded but were relatively few in number and were not subjected to analysis.

Apparatus

For the purpose of this experiment a series of 16 simulated instrument dials was prepared. A sample dial and pointer are shown in Figure 1. Four sizes of dials were used as follows: 1, 1⅞, 2¾, and 4 inches in diameter. The particular dimensions of the two intermediate sizes were chosen to duplicate standard aircraft instruments. Each size of dial was produced with four different graduation intervals, defined in terms of the angular separation between scale marks, as follows: 5, 10, 20, and 40 degrees. Except for the variations in diameter and size of graduation intervals, all dials were identical. The intermediate graduation marks were ⅛ inch in length and approximately 0.02 inches in width. The major graduation marks at each end of the scale were the same width but ¼ inch in length. The numerals on all dials were ⅛ inch in height. All pointers were 3/32 inch in width and of such a length that the tip reached to the inner edge of the shortest graduation marks. All dials covered a range of from 0 to 50 units as shown in Figure 1, with graduation marks only at the 0, 10, 20, 30, 40, and 50 positions, and numerals only at 0 and 50. These dials were engraved on brass plates, which were then painted a flat black and the engraved markings filled with yellow fluorescing paint (pale yellow in daylight) as used on the latest type of USAF instruments.

The experimental dials were presented singly in a panel opening 30 inches from and perpendicular to the subject's eyes. Daylight conditions were simulated with a fluorescent type daylight lamp which provided an illumination of 45 foot-candles at the panel opening. For simulation of night conditions the subject's room was completely darkened and the dial illuminated with a standard C-5 ultra-violet aircraft instrument panel light operating at maximum intensity. No means were available for

obtaining a quantitative measurement of the brightness of the scale markings under ultra-violet illumination. Covering the opening in which the experimental dials were presented was a mechanical shutter operated by the experimenter.

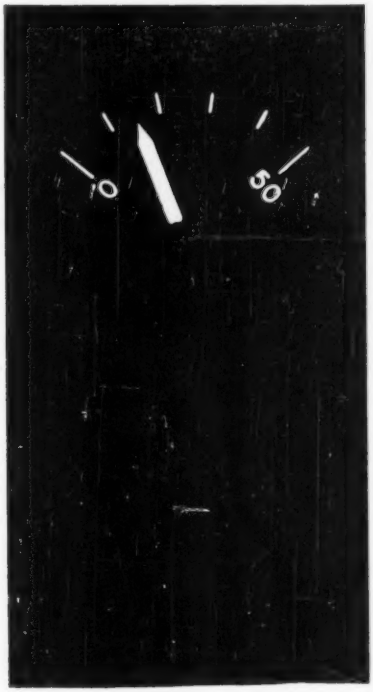


FIG. 1. Sample dial and pointer (1⅞ inch diameter and 20 degree angular separation between scale marks).

On the experimenter's side of the test panel was a carriage on which four of the dials could be mounted side by side. This carriage rode upon two horizontal tracks parallel to the screen. To present any one of the dials the experimenter moved the carriage so that the desired dial would appear in the panel opening. At the experimenter's side of the carriage were four master setting dials 5 inches in diameter. On each of these dials was a pointer connected to the same shaft as the pointer on the dial to be read by the subject. On the experimenter's dials were closely spaced graduations which made possible accurate settings to one-tenth of the space between graduations on the subject's dials.

Also provided at the experimenter's station was a lever for manual operation of the shutter used to expose the dial to the subject. This lever was used also to operate an electric timer through a suitable switch. Thus, the timer indicated the time during which the shutter remained open. Since the experimenter closed the shutter as soon as the dial reading had been completed, the reading on the clock gave a crude measure of the reaction time on each test trial. Several other methods of measuring reaction time were tried but found to be unsatisfactory.

Eighty male college students were used as subjects in this experiment. Only men with 20–20 binocular vision (corrected or uncorrected) were accepted. The subjects were seated in a chair in front of the screen with their eyes 30 inches from the panel opening and with the line of sight perpendicular to the panel opening in order to eliminate parallax. The subjects were divided into groups of 20, each group being tested on a set of four dials. The four dials included one of each diameter and one of each graduation interval. Each subject was given a total of 80 trials, equally divided among the four dials in a random sequence. Of each group of 20 subjects, 10 were tested under simulated daylight conditions and the remaining 10 under simulated night conditions.

A variety of dial settings were chosen so as to represent all portions of the dial from 0 to 50. The actual numbers to be read were the same for all dials although the order of presentation was randomized. The subjects were instructed to read the dials as quickly and accurately as possible to the nearest whole number. As can be seen in Figure 1, the reading to the nearest whole number required estimation to the nearest one-tenth of the distance between graduations.

On each trial the experimenter set the pointer of the dial to be presented, then opened the shutter and waited for the subject's verbal response, following which the shutter was closed and the subject's reading and the clock score recorded.

Results

The experimental design resulted in 200 readings on each of the 16 specific dials under each of the lighting conditions. For each reading both error and time data were obtained. The error data consisted of the deviations of the readings from the actual settings. These deviations could be either negative or positive and increased in step intervals of one, or one-tenth of the space between graduation marks. For purposes of analysis, however, error distributions were made without regard to sign. Since the distributions of these errors, and also response times, were considerably skewed, with the modal error being zero for many of the dials, the statistical treatment presented in this report is limited to medians and 75th percentiles. Means were computed for all the data and found

to present the same general picture as the medians, but the values were inflated because of the skewness.

A summary of the error data for the 16 dials is shown in Table 1. In the third column, it will be noted, the two variables of dial diameter and angular spacing of the divisions have been reduced to a single variable, namely, the length of graduation interval defined as the arc between the inner ends of the shortest scale marks. This value can be described also as the distance the pointer tip must travel between adjacent graduations.

Table 1

Summary of Data on Accuracy of Pointer Interpolation as Function of Dial Diameter and Spacing of Scale Divisions

Dial Diameter, inches	Angular Spacing, degrees	Length of Inner Arc, inches*	Median Error Daylight, % of interval	Median Error Night, % of interval	Median Error Combined, % of interval	75th Percentile Error, Combined, % of interval	Median Error Combined, degrees
1	5	.032	21.8	31.0	26.4	45.5	1.32
1	10	.065	17.8	18.8	18.3	28.5	1.83
1	20	.130	14.3	14.9	14.6	21.2	2.92
1	40	.261	13.2	12.1	12.6	17.6	5.04
1⅞	5	.070	20.0	19.4	19.7	31.8	0.99
1⅞	10	.141	11.2	14.2	12.8	18.5	1.28
1⅞	20	.238	12.4	10.3	11.4	17.1	2.28
1⅞	40	.567	7.8	8.1	8.0	13.9	3.20
2¾	5	.109	15.4	15.7	15.5	21.2	0.78
2¾	10	.218	12.0	9.3	10.6	16.5	1.06
2¾	20	.436	9.4	9.1	9.3	15.2	1.86
2¾	40	.872	8.1	8.1	8.1	14.1	3.24
4	5	.163	14.9	13.9	14.4	20.0	0.72
4	10	.327	9.7	9.0	9.3	15.2	0.93
4	20	.654	8.8	7.9	8.3	14.3	1.66
4	40	1.309	9.8	9.1	9.4	15.0	3.76

* For conversion to minutes of visual angle multiply by 114.7.

This length of the inner arc is presented in inches with the multiplying factor provided for conversion to minutes of visual angle for the 30-inch viewing distance. A comparison of the columns of median errors for daylight and night conditions in Table 1 reveals no consistent difference between these two lighting conditions. Only for the dial with the most closely spaced divisions does the performance under daylight appear to be superior. For this reason it seemed safe to combine the two sets of error data in the remaining columns of the table.

Some of the most important findings contained in Table 1 are presented graphically in Figures 2 and 3. With length of graduation interval along the base line, the median and 75th percentile errors are plotted as per cent of the interval in Figure 2, and as absolute values in Figure 3.

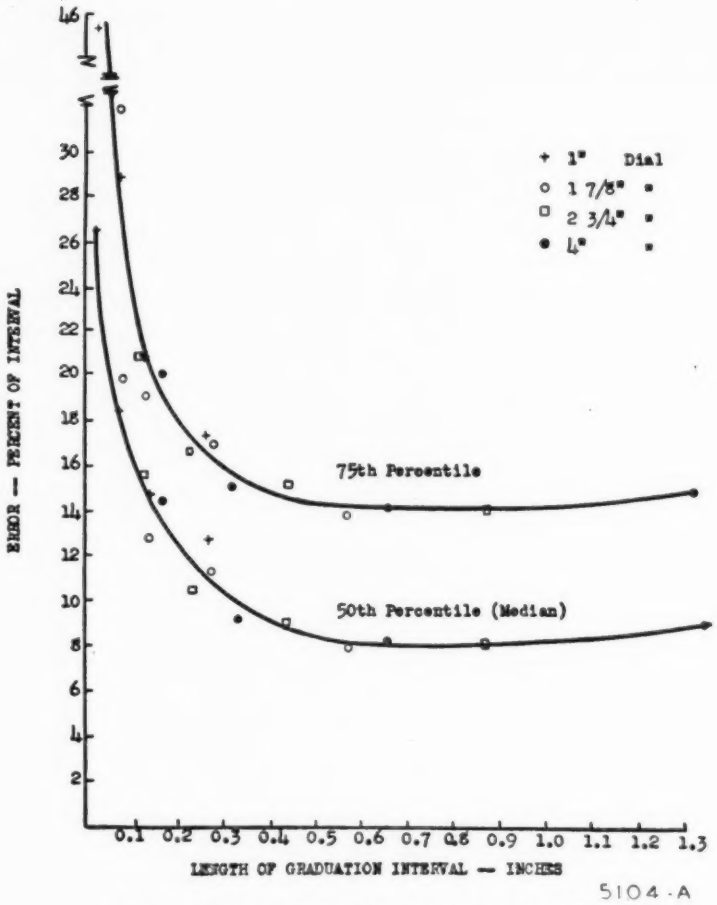


FIG. 2. Relative error of interpolation as a function of length of graduation interval.

Figure 2 will be recognized as a typical Weber function in which threshold ratios (DI/I) are plotted as a function of the stimulus intensity (I). In this figure, it will be noted, the relative accuracy of interpolation is very nearly constant for graduation intervals above 0.5 inch, with a slight rise for the largest interval.

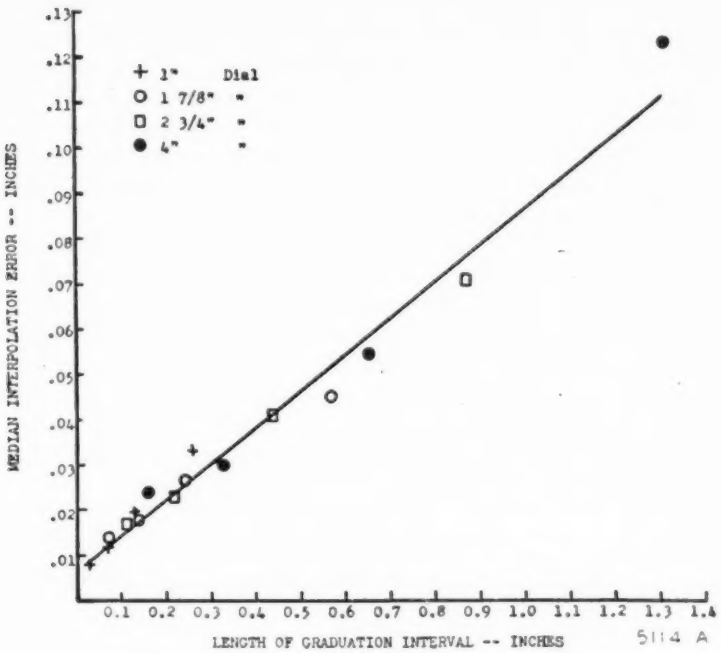


FIG. 3. Absolute error of interpolation as a function of length of graduation interval.

Table 2

Median Time for Interpolation of Pointer Position as a Function of Dial Diameter and Spacing of Scale Divisions

Dial diameter, inches	Graduation Interval 5°	10°	20°	40°
	Seconds per reading for daylight conditions			
1	1.98	1.78	1.91	1.84
1⅞	1.73	1.80	1.86	1.76
2¾	1.83	1.85	1.73	1.77
4	1.87	1.75	1.68	1.90
	Seconds per reading for night conditions			
1	2.48	2.26	2.10	2.13
1⅞	2.18	2.14	2.02	2.10
2¾	2.15	2.05	2.02	2.06
4	2.00	2.02	2.08	2.23

In Figure 3 the absolute values for interpolation thresholds fall very nearly on a straight line, which if extrapolated to zero intervals would intercept the ordinate at approximately 0.006 inch. It is apparent from this curve that no limit had been reached for absolute accuracy of interpolation in this experiment. The excellence with which the various points fit the curves in these two figures indicates that the combination of dial diameter and angular separation into the single variable of length of graduation interval was justified.

The results of the measurements of response time in this experiment are summarized in Table 2. It is apparent that there are no consistent relationships between response times and the dial dimensions, although the method of measuring response time may have been too crude to demonstrate minor relationships that might have been present. On the other hand the response times for the night viewing conditions are consistently higher than under the daylight conditions. It is quite possible that this latter finding was an artifact resulting from a slight delay between opening of the shutter and the fluorescing of the scale marks.

Discussion

Effect of Illumination on Interpolation Accuracy. It is apparent from Table 1 that there was no difference in accuracy of dial reading for day and night conditions except for the dial with the most closely spaced graduations. This finding is in general agreement with results obtained by Spragg and Rock (6) in an investigation of the accuracy of interpolation as a function of illumination. These investigators found accuracy to be almost constant down to a scale mark brightness of 0.022 foot-lambert. In the present experiment the brightness of the scale markings under the simulated night conditions is estimated to have been considerably above the 0.022 foot-lambert value below which Spragg and Rock found a marked loss in accuracy of interpolation.

Effect of Separation between Scale Marks on Speed of Reading. It is noteworthy that in this experiment no relationship was found between the space between graduation marks and speed of reading, although it is admitted that the method of measuring speed of reading lacked precision. In experiments conducted at Princeton University reading time increased as the space between marks was decreased, but in all cases there were changes in other variables which could have caused the changes in reading speed. In the first experiment by Kappauf, Smith, and Bray (3) the reduction in space between scale marks was accompanied by reductions in all other dial dimensions. In the second experiment by Kappauf and Smith (4) reduction of the space between marks was accompanied in some cases by reduction in all other dimensions, in other cases by an increase

in total range of values covered by the scale. The question of whether or not the separation between scale marks in isolation has any effect on speed of dial reading does not appear to have been answered definitely.

Effect of Dial Dimensions on Relative and Absolute Accuracy of Interpolation. In discussing the findings regarding accuracy of interpolation the distinction must be constantly kept in mind between accuracy relative to the interpolation space, and accuracy in absolute units such as degrees or inches. It is apparent from Fig 2 that there is scarcely any useful gain in relative accuracy of interpolation as the graduation intervals are increased beyond 0.25 inch. On the other hand as the intervals are reduced below this value relative accuracy falls off very rapidly. Approximately one-fourth (0.25) to one-half (0.50) inch would therefore seem to be an optimum value for graduation intervals from the standpoint of relative accuracy.

For maximum accuracy in an absolute sense it would appear from Figure 3 that the optimum graduation interval, if there is such, is below the range covered by the present experiment. The data in Figure 3 suggest that the absolute value of interpolation errors might continue to decrease with decreases in graduation interval until the limit of visual acuity is reached. If this is true the limit of visual acuity would determine the optimum graduation interval for maximum accuracy of dial reading. The data of Loucks (5) and Kappauf, Smith, and Bray (3) suggest that as the distance between graduation marks is decreased, there is an increasing tendency to make comprehension errors, that is, assign the wrong values to scale marks. Also, Kappauf and Smith (4) have found that increasing the total number of marks on the scale increases the time required for reading. It would seem, therefore, that there is no easy answer to the problem of what is the optimum interval size for instrument scales, but that the optimum interval will vary with reading criteria.

Summary

Measurements were made of the accuracy of interpolating pointer position between scale marks as a function of dial diameter and the angular spacing between divisions. Subjects were required to estimate the pointer position to within one-tenth the space between graduations. The experimental dials were painted with yellow fluorescing paint on a black background and were read under simulated daylight (45 foot-candles) and night (ultra-violet) illumination conditions. The major results of this investigation may be summarized as follows:

1. Dial diameter and angular spacing of the scale marks could be combined into the single variable of length of graduation interval.

2. The relative error of interpolation decreased as the length of the

graduation interval increased up to approximately 0.5 inch, and was very nearly constant at higher intervals (see Figure 2).

3. The absolute error of interpolation increased very nearly as a linear function of the length of the graduation interval (see Figure 3). If there is an optimum interval for absolute accuracy it would appear to be below the interval lengths used in this study.

4. Except in the case of the most closely spaced divisions the accuracy of interpolation was independent of the two illumination conditions.

5. The speed of dial reading was not systematically related to either dial diameter or angular spacing of the divisions, although the measurements were admittedly crude. Slower reading under the simulated night (ultra-violet) lighting conditions was probably due in part to delay in fluorescence of the dial markings.

Received March 18, 1949.

References

1. Grether, W. F. Factors in the design of clock dials which affect speed and accuracy of readings in the 2400-hour time system. *J. appl. Psychol.*, 1948, **32**, 159–169.
2. Grether, W. F. Psychological factors in instrument reading. I. The design of long-scale indicators for speed and accuracy of quantitative readings. *J. appl. Psychol.*, 1949, **33**, 363–372.
3. Kappauf, W. E., Smith, W. M., and Bray, C. W. *Design of instrument dials for maximum legibility. I. Development of methodology and some preliminary results.* USAF Air Materiel Command Memorandum Report No. TSEAA-694-1L, 20 October 1947.
4. Kappauf, W. E., and Smith, W. M. *Design of instrument dials for maximum legibility. II. A preliminary experiment on dial size and graduation.* USAF Air Materiel Command Memorandum Report No. MCREXD-694-1N, 12 July 1948.
5. Loucks, R. B. *Legibility of aircraft instrument dials: The relative legibility of tachometer dials.* AAF School of Aviation Medicine, Project No. 265, Report No. 1, 1944.
6. Spragg, S. D. S., and Rock, M. L. *Dial reading performance as related to illumination variables. I. Intensity.* USAF Air Materiel Command Memorandum Report No. MCREXD-694-21, 1 October 1948.

Identification of Cola Beverages. III. A Final Study *

N. H. Pronko and J. W. Bowles, Jr.

Wichita University

In a preliminary study, the authors (2) found that when four Cola beverages (Coca Cola, Pepsi Cola, Royal Crown Cola and Vess Cola) were presented to *S*s to identify, the identification responses of those *S*s tended to cluster around the three brand names, Coca Cola, Pepsi Cola and RC Cola, the fourth sample being consistently labelled with one of the three names previously employed. This occurred both when *S*s were given four samples each of *different* Cola beverages or four samples of the *same* Cola.

It was therefore decided to employ only three different Colas on the hypothesis that since our *S*s apparently did not discriminate the Colas on a gustatory basis, their identification responses would be distributed among the three brands in an order approximating chance. Such a procedure was carried out in Experiment II (1) which showed that whether *S*s were given three different beverages or the same beverage three different times, their identifications were not essentially different in the two cases. It was concluded that when *S*s were asked to identify the three leading brands of Cola, they might just as well have drawn their names out of a hat.

Another experiment suggested itself as the logical outcome. If Coca Cola, Pepsi Cola and RC Cola identification responses were randomly distributed when these beverages were actually given serially or otherwise, what would the distribution of those responses be when some relatively unknown Cola drinks were given *instead* of those three? Therefore, the present experiment utilized three brands of less well known or advertised Cola. These were Hyde Park Cola, Kroger Cola, and Spur Cola. The last beverage is now, for the first time, being distributed in this community although none is as generally available as the three leading brands.

Procedure

As in Experiment II, the present study employed two groups of *S*s, 96 in Part I and 60 in Part II. For the most part, these were beginning students in Elementary Psychology.

* The writers wish to express their appreciation to Mr. Fred Snyder, Mr. Donald Synolds and Mr. Glen Allen for their assistance in the experiment reported here.

Part I. Each of 96 *S*s was admitted individually into the experimental room and was invited to sit down. The following instructions were then read to him.

We would like to have you taste and identify some Cola drinks. You will be told in what order and when you are to drink them. After you have finished each sample, report your identification to E and take enough water from the paper cup to rinse your mouth well.

A tray containing three one-oz. glasses of Hyde Park Cola, Kroger Cola and Spur Cola respectively were placed before the *S*. He was then told to drink the beverages labelled x, y, and z in the order indicated to him. Samplings were spaced about a minute apart, *S*'s name and other information being recorded in the interval between drinks.

Order of presentation of beverages, determined preexperimentally, was such that the three stimuli appeared in the first, second and third position 32 times. Such a counter-balanced order was used in order to preclude the operation of position effects or stimuli interactions orally. All beverages were at all times kept out of sight of *S*s and were placed in a refrigerant maintained at approximately 5° C.

Part II. In part II 60 *S*s were administered *the same* Cola drinks at each of three trials. Thus, 20 were given all Hyde Park Cola; 20, all Kroger Cola; and 20, all Spur Cola.

Results

Table 1 shows the distribution of the identification responses of the 96 *S*s of part I who were given one-oz. samples each of Hyde Park Cola, Kroger Cola and Spur Cola. The most conspicuous finding to be observed here is the total absence of correct identifications! Not on a single occasion were any of these beverages correctly named. As a matter of fact, it will be noted that the greater proportion of identification responses are to be found in the Coca Cola, Pepsi Cola and RC Cola columns with a sprinkling of namings under Seven Up, Dr. Pepper and Cleo Cola.

Table 1

Showing the Distribution of 288 Identification Responses when Each of the 96 Ss was Presented in Turn, but in Counterbalanced Order, with One 1 oz. Sample Each of Hyde Park, Kroger, and Spur Cola

Brand of Beverage Given S	Frequency of *S*s' Identification Responses										
	H.P.	K.	S.	C.C.	Pep.	R.C.	7 up	Dr. Pep.	Cleo.	Other	Total
Hyde Park	0	0	0	39	27	24	1	1	0	4	96
Kroger	0	0	0	30	39	22	1	0	1	3	96
Spur	0	0	0	34	33	22	2	1	1	3	96
Total	0	0	0	103	99	68	4	2	2	10	288

Our *S*s in this experiment show essentially the same behaviors as those of Experiment II who were actually given Coca Cola, Pepsi Cola and RC Cola. Our hypothesis is, therefore, substantiated and we must conclude that our *S*s have not discriminated from among the varieties of Cola beverages employed in our series of studies. Instead, they have applied a readily available repertoire of naming reactions whose probable source is advertising, familiarity through actual contact and other forms of culturization. One more point should be noted. In the total of 288 identification responses, observe that the three relatively less well known beverages here employed were identified as Coca Cola 103 times, as Pepsi Cola 99, and as RC Cola 68 times. Again, as in previous studies, Coca Cola and Pepsi Cola names are applied with greatest frequency, with RC trailing as before.

Table 2

Showing the Distribution of 180 Identification Responses when Each of 60 Ss was Presented with Three 1 oz. Glasses of Either the Single Brand, Hyde Park, Kroger, or Spur Cola

Brand of Beverage Given S	Frequency of Ss' Identification Responses										
	H.P.	K.	S.	C.C.	Pep.	R.C.	7 up	Dr. Pep.	Cleo.	Other	Total
Hyde Park	0	0	0	25	21	12	0	1	1	0	60
Kroger	0	0	0	22	25	13	0	0	0	0	60
Spur	0	0	0	26	19	9	0	1	2	3	60
Total	0	0	0	73	65	34	0	2	3	3	180

Table 2 shows distribution of the 180 identifications of our 60 *S*s who were presented with three samples of the *same* Cola. The reader will note that whether our *S*s get three samples of Hyde Park, Kroger or Spur Cola, they nevertheless identify each of them most frequently as Coca Cola and Pepsi Cola and less frequently as RC Cola with the same sprinkling of other brands as before. The pattern of total frequencies is the same as in all parts of this and other experiments with Coca Cola and Pepsi Cola in the lead and RC as runner-up. We suggest that this consistent pattern may reflect the relative efficacy of the advertising of these three brands.

Summary

A group of 96 *S*s was asked to identify one-oz. samples of Hyde Park Cola, Kroger Cola and Spur Cola presented in counterbalanced order.

1. There were *no* correct identifications. Instead, these beverages were identified 103 times as Coca Cola, 99 times as Pepsi Cola, 68 times as RC and a total of 18 times as some other soft drink.

2. Another group of 60 *S*s, each of which was given three one-oz. samples of only one of the three Colas, showed an assortment of Coca Cola (73), Pepsi Cola (65) and RC Cola (34) identifications, again indicating total absence of correct judgments.

3. It is concluded that no matter whether our *S*s are asked to identify three or four of the *same* or *different* Cola beverages, regardless of what particular brands are employed, their identification responses are inevitably "Cola Cola, Pepsi Cola, RC Cola."

4. The seven brands of Cola beverages employed in our series of studies appear to have the same stimulus function for our *S*s and may be said to be "equivalent stimuli."

Received June 14, 1948.

References

1. Bowles, J. W., Jr., and Pronko, N. H. Identification of Cola Beverages. II. A further study. *J. appl. Psychol.*, 1948, **32**, 559–564.
2. Pronko, N. H., and Bowles, J. W., Jr. Identification of Cola beverages. I. First study. *J. appl. Psychol.*, 1948, **32**, 304–312.

Book Reviews

Samuel A. Stouffer et al. *The American soldier: Volume I, Adjustment during army life; Volume II, Combat and its aftermath.* Princeton: Princeton University Press, 1949. Pp. 600 each. Vol. 1 and 2, $13.50. Separate, $7.50.

These are the first two of four volumes prepared and edited under the auspices of a Special Committee of the Social Science Research Council. The material reported is based upon data concerning soldiers' attitudes collected by the Research Branch of the Information and Education Division, War Department, during the period, December, 1941, to August, 1945.

The basic tool which the Research Branch used was the questionnaire survey. A typical survey went through the following stages: (1) a request for information concerning attitudes or behavior of soldiers by some agency or branch of the army; (2) conferences between members of the Research Branch and the requesting agency in which the area to be investigated was outlined; (3) a field trip by two or three members of the Research Branch in which casual conversations were held with enlisted men and officers on the topics to be covered in the survey; (4) preparation of a questionnaire based upon the experience gained in the field trip; (5) a pre-test of the questionnaire, followed by revisions; (6) final conferences with the requesting agency on the revised questionnaire; (7) clearance of the questionnaire with the Director of the Research Branch; (8) sending the questionnaire into the field.

Samples for domestic surveys were selected in Washington; those for overseas surveys were selected in theater headquarters. In general the questionnaires were filled out directly by the men, personal interviews being limited to those in the lower educational groups.

Over 200 such surveys were made by the Research Branch in which practices, preferences, and attitudes were investigated. The following partial list indicates something of the diversity of the areas surveyed: use of atabrine; practices related to trench foot; housing preferences of troops stationed in Alaska; popularity of various departments in *Yank;* radio listening habits; attitudes toward the British, Chinese, and Germans; leisure-time activities; attitudes toward the WACS, civilians, Negroes, and MP's; practices in relation to venereal disease prevention; attitudes toward medical care, hospitalization, war, savings and insurance plans, rotation and demobilization; reactions to army films, orientation and

indoctrination programs; attitudes toward service in the tropics, combat duty, job assignment, and training; reactions to German weapons; desires for educational courses; preferences in Christmas gifts; and studies of map reading, word pronunciation, and psychoneuroses.

Attitudes of special groups on various problems were surveyed. Some of the groups investigated were: Negroes, combat troops, Special Service officers, MP's, WACS, psychiatric patients, combat infantrymen, paratroopers, AWOL's, hospital nurses, combat veterans, Air Force returnees, combat veterans in hospitals, B-29 officers and enlisted men, and hospital patients.

As a result of these surveys, some concrete actions were taken by the army. One example: the point system for demobilization was based upon a survey of the factors which enlisted men thought should be considered and the relative weights which they thought should be assigned to these factors. In addition, a monthly bulletin, *What the Soldier Thinks*, was published and circulated down to regimental commanders. Another bulletin, the *Monthly Progress Report*, was prepared for higher level officers. Both bulletins were intended to keep officers informed concerning the attitudes of enlisted men and officers on various problems of interest to the army.

Although the first two volumes are primarily intended for the army, historians, social psychologists, and sociologists, they contain much useful information for the industrial psychologist and personnel workers. This is indicated by the following topics discussed in some detail in the various chapters: job satisfaction and job morale; leadership and social control; motivation; social mobility; adjustment; promotion; training. Most of these subjects are covered in Volume I which deals with the period of adjustment during army life. Volume II is primarily concerned with combat and its aftermath. Since many readers of this journal will be primarily interested in Volume I, it is unfortunate that separate indexes for the two volumes were not prepared. As it is, a single index is included in Volume II.

It is worth mentioning also that the index does not contain entries for the pertinent problems of *reliability* and *validity*, although some attempts at validation of the questionnaires were made and are discussed in the volumes. Another point of minor irritation is the frequent forward reference in Volumes I and II to the material to be published in Volumes III and IV.

Charts, graphs, and tables are to be found on almost every other page. Some of these are very elaborate and detailed. Others show the percentage of various groups responding to a single question in a particular manner. The basic statistic is the percentage, although a few correlation co-

efficients are reported and an occasional chi square is to be found—usually in a footnote. This is not intended as a criticism. The reader should keep in mind that the work of the Research Branch was primarily directed toward the collection of information about attitudes which would be useful to the army planners of information, orientation, and educational programs. Their job was, as one of the authors calls it, a "practical engineering" job. From the evidence reported in these two volumes, it was a job well done.

Allen L. Edwards

The University of Washington

Planty, Earl G., McCord, William S., and Efferson, Carlos A. *Training employees and managers for production and teamwork.* New York: Ronald Press, 1948, pp xiii + 278. $5.00.

This book is characterized by breadth of viewpoint, purpose, scope, and principles. Training is conceived as a training of the whole employee (on any job level) in attitudes, skills, and knowledges in such way that he fits into the framework of his job, his company, and the American system of free enterprise. The ambitious program undertaken by the authors could easily have led to a superficial or biased book. For the most part these potential dangers have been avoided. The authors indicate considerable evidence of sound professional training, practical experience, common sense, and penetrating insight. The result is a book which will be of considerable value and interest to all persons directly or indirectly engaged in training in business and industry.

The first two chapters cover what training is and what it will do. The second part, consisting of the next six chapters, deals with organizing, installing, and administering a training program. Basic principles are given together with specific techniques and mechanics. The third part of the book contains thirteen how-to-do-it chapters devoted to training programs and methods. Chapters are included on teaching aids, teaching in business and industry, the training staff, the small company, helpful resources, and training for specific groups such as new employees, supervisors, technical and professional, trade and semiskill, and office employees.

It was difficult for this reviewer to remain objective while appraising the book. The enthusiasm aroused by the first few chapters kept increasing throughout the entire book except for the chapter on technical and professional training. This chapter is of little value to the book or the reader except for the implication that such training has lagged far behind other types. A serious error of omission is the failure of the book to discuss training costs. Such a chapter would appear imperative in a book of this type.

The authors have shown good judgment in striking a balance between various topics, have organized the material excellently, and have used well chosen examples and illustrations. They are to be commended for attacking some of the cliches currently held by training men; as, for example, the universal superiority of conferences as compared with lectures. The choice of words and mechanics of writing are considerably better than usually found in books of this type. These factors all add up to the best book on industrial training which this reviewer has ever seen.

Clifford E. Jurgensen

Minneapolis Gas Company

deFord, Miriam Allen. *Psychologist unretired, the life pattern of Lillian Martin.* California: Stanford University Press, 1948. Pp. 130. $3.00.

Lillian Martin, a most versatile psychologist, came from a distinguished family, was self-supporting in her education, took her undergraduate work at Vassar, spent four years in graduate study of psychology in Germany in the 90's, held a professorship of psychology in Stanford University from which she retired for age at 65 in 1916. She then turned to private practice and hospital service as a clinical psychologist and worked enthusiastically and with notable success for 27 years until she died at the age of 91.

This book will be welcomed by lovers of biography as a model of literary style, by all clinical psychologists, and by all students of guidance in the art of aging.

Carl E. Seashore

Winter Park, Florida

Kaback, Goldie Ruth. *Vocational personalities: an application of the Rorschach Group Method.* New York, Columbia University: Bureau of Publications, 1946. Teachers College Contributions to Education No. 924. Pp. xii + 116. $2.10.

This study tests the hypothesis that vocational choice is in part a function of personality, the differences in personality responsible for a given occupational choice being measurable by the Rorschach Group Method.

Group Rorschachs were obtained from 75 accountants, 75 pharmacists, 75 accountancy students and 75 pharmacy students. Brief life data sheets were filled in by the subjects. Mean ages were respectively 37, 37, 18, and 20 years, ranges being greater for the pharmacist groups than for their comparison accountant groups. Accountants averaged more education than pharmacists; students were similar to each other. Pharmacists had 16, accountants 13 years of work experience in their field. A miscellaneous vocational group numbering 108 was obtained from (ap-

parently) a co-worker and responses were compared with the four criterion groups studied. Reliability of Rorschach scoring was not reported.

Kaback finds that she can differentiate accountants from pharmacists by multiple correlation of .54 between the criterion and 24 Rorschach components. This *r* is too low for effective prediction. Accountants were significantly higher than pharmacists in W, d, R, P, O, Fc, M, FM, Fm, H+A, Aobj., Pl and Obj. Pharmacists were significantly higher on Fk and At. This means (roughly) that accountants were more productive (this factor alone seems to account for many of the differences shown) higher intellectually, more "socially sensitive—or aware of a need for social contact," more cautious, creative and original and perhaps more widely diversified in interests than pharmacists, while the latter showed more "anxiety." These differences, as the author points out, while statistically significant, are not of great practical value in vocational counseling.

For students, a multiple *r* against the criterion of .653 was shown (a respectable *r*, but not high enough for good prediction.) Accountants in training were higher than pharmacists in training on R, P, H+A, W, M, Fm, F, FC, Obj., D, Emb., and Arch; hence differences resemble the professional differences.

The author concludes: all groups normal, reasonably well adjusted (accountancy students best), accountants more intelligent, pharmacists less mature and impulsive (for both students and professional persons); and students of both groups show selves as more practical than the professional groups.

The author is to be commended for a type of study greatly needed. It seems however, first, that such a study can hardly attain highly reliable or valid results as long as socioeconomic pressures, family predilections, etc., determine the chosen career, often quite without regard to the personal wishes or capabilities of the individual involved; and that, second, a Rorschach study is difficult if not impossible when scoring categories of the test are considered (as is almost necessary) without the benefit of refinements in statistics of percentages, ratios, and sub-relationships.

The most ardent proponent of the Rorschach would probably not regard it as infallible when adapted as a group method and scored for means, using isolated scoring categories. In the hands of a competent technician, used to illuminate other data, it often seems to be of great use in vocational counseling; more "clinical" validations along the rather promising lines shown by the author may result from this interesting, carefully done, but practically not too useful study, the final conclusion of which is necessarily that there are all types of persons who enter or study in the fields of accountancy and pharmacy.

Boyd McCandless

The Ohio State University

Eysenck, H. J. *Dimensions of personality.* Cambridge, at the University Press; New York; The Macmillan Company, 1947. Pp. 308. $5.00.

This book is the result of a cooperative effort to discover the main dimensions of personality and to define them operationally by means of strict experimental, quantitative procedures. About forty distinct researches were carried out on some 10,000 normal and neurotic subjects by a research team of psychologists and psychiatrists at the Mill Hill Emergency Hospital, a war time neurosis center in London. Dr. Eysenck and his collaborators have turned to fruitful purpose the opportunities such a place affords for studying some of the main factors of personality and have shown how profitable may be the collaboration of psychologists and psychiatrists in pursuit of this aim.

The first chapter entitled, *Methods and Definitions* describes the methodological conditions underlying the researches, and the working concepts and theories of temperament and personality structure. The second chapter, entitled, *Assessments and Ratings* presents a factorial analysis of 39 trait ratings, carried out by psychiatrists on 700 neurotic patients; also described is a study with questionnaires dealing with "neuroticism," "persistence," and "irritability." *Physique and Constitution* concerns studies performed on normal and neurotic subjects through the methods of factorial analysis wherein main dimensions of body configuration were isolated objectively; certain personality differences were also found with respect to body size. The fourth chapter, *Ability and Efficiency* presents discussions and data on intelligence and neurosis, "scatter," level of aspiration, perseveration, and persistence. *Suggestibility and Hypnosis* attempts to distinguish various types of suggestibility, to establish the relation between suggestibility and hypnosis, and to discover personality correlates of suggestibility (its relation to hysteria and neuroticism). Chapter six, *Appreciation and Expression* offers experimental evidence on preference judgments. *Synthesis and Conclusions*, chapter seven, summarizes the researches indicating two main personality dimensions namely, "neuroticism" and "extraversion-introversion," wherein the former is a general factor in the conative sphere, while the latter is a general factor in the affective sphere.

While this book may presently contribute little that is useful to harrassed personnel psychologists, these specialists should nevertheless be familiar with the methods of investigation utilized. Much that is presented may influence our concepts of psychopathology in the future. However, it is recommended to all applied psychologists working in clinical fields and is "must" reading for the small corps of experimental clinical psychologists.

Arthur Weider

University of Louisville School of Medicine,
Dept. of Psychiatry, Div. of Psychosomatic Medicine

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

The Harvard list of books in psychology. Gordon W. Allport et al. Cambridge: Harvard University Press, 1949. Pp. 77. $1.00.

Differential psychology. Revised edition. Anne Anastasi and John P. Foley, Jr. New York: The Macmillan Co., 1949. Pp. 894. $5.50.

Twilight in India. Gervee Baronte. New York: Philosophical Library, 1949. Pp. 382. $3.75.

Orientation et selection professionnelles par l'examen psychologique du caractere. Fr. Baumgarten. Paris, France: Dunod, Editeur, 92, Rue Bonaparte, 1949. Pp. 184. 680 fr.

Social psychology of modern life. Revised edition. Steuart H. Britt. New York: Rinehart and Co., Inc., 1949. Pp. 703. $4.50.

Labor relations and hindrances to full production. S. L. Burk. San Francisco: California Personnel Management Association, 1948. Pp. 18. $1.00.

The neurosis of man. Trigant Burrow. New York: Harcourt, Brace, 1949. Pp. 428. $7.50.

Reading manual and workbook. Homer L. J. Carter and Dorothy J. McGinnis. New York: Prentice-Hall, Inc., 1949. Pp. 120. $1.75.

Human relations in public administration. Alfred De Grazia. Chicago: Public Administration Service, 1949. Pp. 52. $1.50.

Criteria for the life history. John Dollard. Reprint. New York: Peter Smith, Publisher, 1949. Pp. 294. $3.25.

The supervisor's management guide. M. Joseph Dooher and Vivienne Marquis, Editors. New York: American Management Association, 1949. Pp. 190. $3.50.

We human chemicals. Thomas Dreier. Scarsdale: Updegraff Press, Ltd., 1949. Pp. 122. $2.00.

Mental testing. Florence L. Goodenough. New York: Rinehart and Co., Inc., 1949. Pp. 609. $5.00.

Readings in social security. William Haber and Wilbur Cohen. New York: Prentice-Hall, Inc., 1948. Pp. 634. $5.75.

Psychosexual development in health and disease. Paul H. Hoch and Joseph Zubin, Editors. New York: Grune and Stratton, 1949. Pp. 283. $4.50.

Group guidance, principles, techniques, and evaluation. Robert Hoppock. New York: McGraw-Hill Book Co., Inc., 1949. Pp. 393. $3.75.

Motor performance and growth. A developmental study of static dynamometric growth. Harold E. Jones. Berkeley: University of California Press, 1949. Pp. 181. $3.00, cloth, $2.00, paper.

Helping students find employment. Forrest H. Kirkpatrick et al. Washington, D. C.: American Council on Education Studies, 1949. Pp. 37. $.75.

Studies in human behavior. Merle Lawrence. Princeton: Princeton University Press, 1949. Pp. 184. $3.50.

Explorations in personal adjustment—a workbook. George F. J. Lehner. New York: Prentice-Hall, Inc., 1949. Pp. 144. $1.50.

Mental hygiene in public health. Paul V. Lemkau. New York: McGraw-Hill Book Co., Inc., 1949. Pp. 396. $4.50.

The psychology of personal adjustment. Second edition. Fred McKinney. New York: Wiley and Sons, Inc., 1949. Pp. 752. $6.00.

Selling performance and contentment in relation to school background. Albert C. Mossin. New York: Bureau of Publications, Teachers College, Columbia University, 1949. Pp. 166. $2.75.

The pre-election polls of 1948. Frederick Mosteller et al. New York: Social Science Research Council, 1949. Pp. 396. Paper, $2.50; Cloth, $3.00.

The nature-nurture controversy. Nicholas Pastore. New York: Columbia University Press, 1949. Pp. 213. $3.25.

Perception of symbol orientation and early reading success. Muriel Catherine Potter. New York: Bureau of Publications, Teachers College, Columbia University, 1949. Pp. 69. $2.10.

Human relations in an expanding company. F. L. W. Richardson and Charles R. Walker. New Haven: Yale University Labor and Management Center, 1948. Pp. 95. $1.50.

Letters to my son. Dagobert D. Runes. New York: The Philosophical Library, 1949. Pp. 92. $2.75.

Happiness for husbands and wives. Harold Shryock. Washington, D. C.: Review and Herald Publishing Association, 1949. Pp. 256. $2.50.

Rehabilitation of the handicapped. William H. Soden, Editor. New York: Ronald Press Co., 1949. Pp. 399. $5.00.

Personnel management for supervisors. Claude E. Thompson. New York: Prentice-Hall, Inc., 1949. Pp. 208. $2.95.

Selected writings from a connectionist's psychology. Edward L. Thorndike. New York: Appleton-Century-Crofts, Inc., 1949. Pp. 370. $3.50.

Working with people. Auren Uris and Betty Shapin. New York: The Macmillan Co., 1949. Pp. 314. Probable price, $3.50.

Experimental foundations of general psychology. Third edition. Willard L. Valentine and Delos D. Wickens. New York: Rhinehart and Co., Inc., 1949. Pp. 472. $3.00.

Community under stress. Elizabeth Head Vaughan. Princeton: Princeton University Press, 1949. Pp. 160. $2.50.

Personnel selection in the British forces. Philip E. Vernon and John B. Parry. London: University of London Press Ltd., 1949. Pp. 324. 20/-net.

Children with mental and physical handicaps. J. E. Wallace Wallin. New York: Prentice-Hall Inc., 1949. Pp. 549. $5.00.

Out-of-school vocational guidance. Roswell Ward. New York: Harper and Brothers, 1949. Pp. 155. $2.50.

Constructing classroom examinations. Ellis Weitzman and Walter J. McNamara. Chicago: Science Research Associates, 1949. Pp. 153. $3.00.

Supervisory training—why, what, how. ILIR Publications, Series A, Vol. 3, No. 3, Urbana: Institute of Labor and Industrial Relations, University of Illinois, 1949. Pp. 24. Gratis.

Recruitment, selection, and indoctrination of female clerical employees. Personnel and Training Series Research Project Report No. 4. New York: Scott Foresman and Co., Miss Edith Harper, 1949. Pp. 35. $2.00.

Sales manager's handbook. Chicago: The Dartnell Corporation, 1949. Pp. 1150. $10.00.

Training for tomorrow: proceedings of the Third Annual Training Conference of Educational Directors in Industry and Commerce. Montreal: Canadian Industrial Trainers' Association, 1949. Pp. 123. $2.00.

Executive personality and job success. New York: American Management Association, 1948. Pp. 35. $.75.

Building worker interest in production problems. New York: American Management Association, 1949. Pp. 34. $.75.

Management's role in industrial mobilization. New York: American Management Association, 1948. Pp. 27. $.50.

Organization controls and executive compensation. New York: American Management Association, 1948. Pp. 54. $1.00.

Personnel functions and the line organization. New York: American Management Association, 1948. Pp. 31. $.75.

Worker morale and productivity. New York: American Management Association, 1948. Pp. 38. $.75.

REPRINTS AND ARTICLES

of the

American Psychological Association

1515 Massachusetts Ave., Northwest, Washington 5, D. C.

Johnny Rocco, by Jean Evans. Second reprinting. From the Journal of Abnormal and Social Psychology, July, 1948. One for 25¢; 50 for $10.00.

A New Readability Yardstick, by Rudolf Flesch. Second reprinting. From the Journal of Applied Psychology, June, 1948. One for 25¢; 50 for $5.00.

Tables for Use with the Flesch Readability Formulas, by J. N. Farr and James J. Jenkins. From the Journal of Applied Psychology, June, 1949. One for 25¢; 50 for $10.00.

Stipends for Graduate Students in Psychology: 1949–50, by the APA office staff. From the American Psychologist, January, 1949. One for 25¢.

Available Internships in Psychology, by Helen Wolfle. From the American Psychologist, February, 1949. One for 25¢.

Occupations in Psychology, by Carroll L. Shartle. The December 1946 issue of the American Psychologist. One for 50¢; 100 for $35.00.

The Psychological and Social Sciences in the National Military Establishment, by Lyle H. Lanier. The May 1949 issue of the American Psychologist. One for 75¢.

Subscription Lists of the American Psychological Association

MEMBERS AND AFFILIATES

Approximately 9,000 names

The American Psychological Association maintains an address list of its members and affiliates, which is for sale providing the nature of its use is in conformity with the purposes of the Association.

	1950 Prices
Envelopes addressed ..... (advertiser furnishes envelopes and pays express charges)	$35.00
Addresses on tape, not gummed ..... (suitable for a mailing machine)	$17.50

STATE LISTS

Priced according to number of names wanted

SUPPLEMENTARY LISTS

Approximately 3,500 names in total list
Individual journal lists vary from 400 to 1,600

The Association also maintains a list of subscribers who are not members of the Association (universities, libraries, industrial laboratories, hospitals, other types of institutions, and individual subscribers). The general list for all journals includes all types. Each single journal has a more specialized circulation.

For any one journal, envelopes addressed	$15.00
For any one journal, addresses on tape	$10.00
For all journals, envelopes addressed	$20.00
For all journals, addresses on tape	$15.00

For further information, write to

American Psychological Association
1515 Massachusetts Avenue Northwest
Washington 5, D. C.

JOURNAL OF APPLIED PSYCHOLOGY

YEAR	VOLUME	AVAILABLE NUMBERS						PRICE PER NUMBER	PRICE PER VOLUME
		MAR	JUN	SEP	DEC				
1917	1	–	–	3	4			$1.75	$3.50
1918	2	1	2	3	4			$1.75	$6.00
1919	3	1	2	3	4			$1.75	$6.00
1920	4	1	2 &	3	4			$1.75	$6.00
1921	5	1	2	3	4			$1.75	$6.00
1922	6	1	2	–	4			$1.75	$5.25
1923	7	–	–	3	4			$1.75	$3.50
1924	8	1	2	3	4			$1.75	$6.00
1925	9	1	2	3	4			$1.75	$6.00
1926	10	–	2	–	–			$1.75	$1.75
		FEB	APR	JUN	AUG	OCT	DEC		
1927	11	1	2	3	4	5	6	$1.25	$6.00
1928	12	1	2	3	4	5	6	$1.25	$6.00
1929	13	1	2	3	4	5	6	$1.25	$6.00
1930	14	1	2	3	4	5	6	$1.25	$6.00
1931	15	1	2	3	4	5	6	$1.25	$6.00
1932	16	1	2	3	4	5	6	$1.25	$6.00
1933	17	1	2	3	4	5	6	$1.25	$6.00
1934	18	1	2	3	4	5	6	$1.25	$6.00
1935	19	1	2	3	4	5	6	$1.25	$6.00
1936	20	1	2	3	4	–	6	$1.25	$6.00
1937	21	1	2	3	4	5	6	$1.25	$6.00
1938	22	1	2	3	4	5	6	$1.25	$6.00
1939	23	1	–	3	4	5	6	$1.25	$6.00
1940	24	–	–	–	4	5	6	$1.25	$3.75
1941	25	1	2	3	4	5	6	$1.25	$6.00
1942	26	1	2	3	4	5	6	$1.25	$6.00
1943	27	1	2	3	4	5	6	$1.25	$6.00
1944	28	1	2	3	4	5	6	$1.25	$6.00
1945	29	1	2	3	4	5	6	$1.25	$6.00
1946	30	1	2	3	4	5	6	$1.25	$6.00
1947	31	–	2	3	4	5	6	$1.25	$6.00
1948	32	1	2	3	4	5	6	$1.25	$6.00
1949	33	By subscription, $6.00						$1.25	—
		List price, Volumes 1 through 32							$179.75
		30% Discount							53.93
		Net price, Volumes 1 through 32							$125.82

Information about the Journal of Applied Psychology: only 14 numbers are out of print in the thirty volumes. These are scattered throughout the years.

Volumes 1 through 10 had four numbers per year; after that there are six numbers per year. Number 3 of Volume 24 will be sold only with an order for several volumes.

Information about prices: the journal is uniformly priced at $6.00 per volume. When four numbers per volume were issued, the price per number is $1.75; when six numbers per year were issued, the price per number is $1.25. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over

20% on orders of $100.00 and over

30% on orders of $150.00 and over

Current subscriptions and orders for back numbers should be addressed to:

AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 Massachusetts Avenue, N. W.

Washington 5, D. C.

Future titles in the series of

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 63, 1949

FACIAL EXPRESSIONS OF EMOTION. James C. Coleman, University of Southern California. *#296, $1.00*

A COMPARATIVE STUDY OF THE WHERRY-DOOLITTLE AND A MULTIPLE CUTTING-SCORE METHOD. Glen Grimsley, General Motors Institute. *#297, $.75*

FACTOR ANALYSES OF TESTS AND CRITERIA: A COMPARATIVE STUDY OF TWO AAF PILOT POPULATIONS. William B. Michael, Princeton University. *#298, $1.00*

THE APPRAISAL OF PARENT BEHAVIOR. Alfred L. Baldwin, Joan Kalhorn, and Fay Huffman Breese, Fels Research Institute. *#299, $1.50*

STUDIES OF IDENTICAL TWINS REARED APART. The late Barbara S. Burks; and Anne Roe, New York City. *#300, $1.00*

COLOR PREFERENCES OF PSYCHIATRIC GROUPS. Samuel J. Warner, New York City. *#301, $.75*

PERCEPTION OF BODY POSITION AND OF THE POSITION OF THE VISUAL FIELD. H. A. Witkin, Brooklyn College. *#302, $1.00*

AN EXPERIMENTAL EXAMINATION OF THE THEMATIC APPERCEPTION TECHNIQUE IN CLINICAL DIAGNOSIS. A. A. Hartman, Boston University. *#303, $1.00*

RELIGION AND HUMANITARIANISM: A STUDY OF INSTITUTIONAL IMPLICATIONS. Clifford Kirkpatrick, Indiana University. *#304, $.75*

THE DEVELOPMENT AND VALIDATION OF A SET OF MUSICAL ABILITY TESTS. Robert W. Lundin, Hamilton College. *#305, $1.00*

A COMPARISON OF TWO TESTS OF INTELLIGENCE ADMINISTERED TO ADULTS. Anna S. Elonen, University of Chicago. *#306, $1.00*

The 1949 volume of the Psychological Monographs will comprise eleven separate issues. We will be glad to place orders for any of these; the orders will be filled when the issues appear. The entire volume may be subscribed to for $6.00. The actual date of publication is uncertain. The 1949 volume will probably be completed by April of 1950.

American Psychological Association

1515 Massachusetts Avenue N. W., Washington 5, D. C.

Just Published—

1949 DIRECTORY

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 MASSACHUSETTS AVENUE N. W.
WASHINGTON 5, D. C.

In the alphabetical list of 6735 members, the 1949 Directory of the Association gives the names of the members, their addresses, their present positions, their last degrees, and their class of membership. Membership lists for the Divisions of the Association, the lists of Diplomates in the fields of clinical, industrial, and counseling of the American Board of Examiners in Professional Psychology, the By-Laws, and a geographical and institutional index of members are included. The editor is Helen M. Wolfle of the Association staff. *250 pages, $3.00.*

SAMPLE ENTRIES

Humphreys, Lloyd G. School of Education, Stanford Univ, Stanford, Calif. Assoc. prof. educ. and psych. PhD 38. A 5; F 9, 19.

Hunnicker, Mr. Albert L. Committee on Human Development, Univ. of Chicago, Chicago, 37, Ill. Stud. MA 39. A.

Hunt, Dr. Howard F. Dept. Psych, Univ. of Chicago 37, Ill. Assoc. prof. PhD 43. F 12.

Hunt, Dr. J. McV. Institute of Welfare Research, Community Service Society, 105 East 22nd St, New York 10, N. Y. Dir. PhD 33. Dipl-Cl. F 3, 8, 9, 12.

Hunt, Mary Louise 1252 Talbert St. S. E, Washington, D. C. A '48.

Hunt, Dr. Thelma Dept. Psych, George Washington Univ, Washington 6, D. C. F 5, 12.

Hunt, Prof. William A. Dept. Psych, Northwestern Univ, Evanston, Ill. Pof. psych. and biol. PhD 31. Dipl-Cl. F 2, 3, 8, 9, 12, 19, 20.

Hunt, Mr. Wilson L. Boston State Hosp, 591 Morton St, Dorchester Center 24, Mass. Clin. psych't. AM 47. A '49.

Hunter, Dr. Elwood C. Dept. Education, Tulane Univ, New Orleans 15, La. Head of dept. PhD 35. A 5, 15.

JOURNAL OF APPLIED PSYCHOLOGY

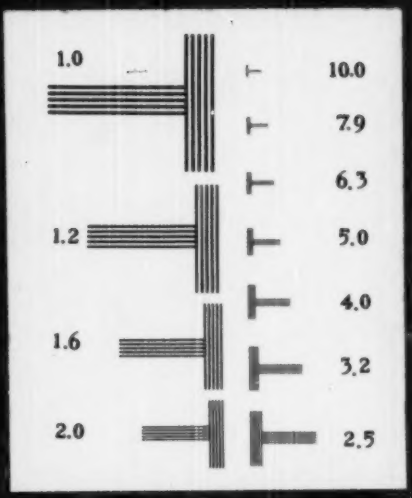
VOL. 34
NOs 1-6

1950

PUB. 277

UNIVERSITY MICROFILMS
ANN ARBOR, MICHIGAN. 1951

RESOLUTION CHART



100 MILLIMETERS

INSTRUCTIONS Resolution is expressed in terms of the lines per millimeter recorded by a particular film under specified conditions. Numerals in chart indicate the number of lines per millimeter in adjacent "T-shaped" groupings.

In microfilming, it is necessary to determine the reduction ratio and multiply the number of lines in the chart by this value to find the number of lines recorded by the film. As an aid in determining the reduction ratio, the line above is 100 millimeters in length. Measuring this line in the film image and dividing the length into 100 gives the reduction ratio. Example: the line is 20 mm. long in the film image, and 100/20 = 5.

Examine "T-shaped" line groupings in the film with microscope, and note the number adjacent to finest lines recorded sharply and distinctly. Multiply this number by the reduction factor to obtain resolving power in lines per millimeter. Example: 7.9 group of lines is clearly recorded while lines in the 10.0 group are not distinctly separated. Reduction ratio is 5, and 7.9 x 5 = 39.5 lines per millimeter recorded satisfactorily. 10.0 x 5 = 50 lines per millimeter which are not recorded satisfactorily. Under the particular conditions, maximum resolution is between 39.5 and 50 lines per millimeter.

Resolution, as measured on the film, is a test of the entire photographic system, including lens, exposure, processing, and other factors. These rarely utilize maximum resolution of the film. Vibrations during exposure, lack of critical focus, and exposures yielding very dense negatives are to be avoided.

THIS PUBLICATION IS REPRODUCED BY AGREEMENT WITH THE COPYRIGHT OWNER. EXTENSIVE DUPLICATION OR RESALE WITHOUT PERMISSION IS PROHIBITED.